

日本国特許庁
JAPAN PATENT OFFICE

501 P1397 US 03
2/14/02
Jc979 U.S. PTO
09/942896
08/30/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application:

2000年 8月31日

出願番号

Application Number:

特願2000-264035

出願人

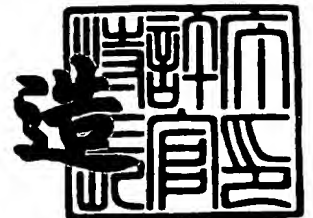
Applicant(s):

ソニー株式会社

2001年 5月25日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



出証番号 出証特2001-3044812

【書類名】 特許願

【整理番号】 0000645904

【提出日】 平成12年 8月31日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/18
G10L 15/06
G10L 15/20

【発明者】

【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社
内

【氏名】 中塚 洪長

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100082131

【弁理士】

【氏名又は名称】 稲本 義雄

【電話番号】 03-3369-6479

【手数料の表示】

【予納台帳番号】 032089

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9708842

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 モデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置

【特許請求の範囲】

【請求項 1】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識に用いられる前記モデルの適応を行うモデル適応装置であって、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出手段と、

前記所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、前記所定のモデルの適応を行うモデル適応手段と

を備えることを特徴とするモデル適応装置。

【請求項 2】 前記モデル適応手段は、前記抽出データの新鮮さを表す新鮮度も用いて、モデルの適応を行う

ことを特徴とする請求項 1 に記載のモデル適応装置。

【請求項 3】 前記モデル適応手段は、前記所定区間における抽出データの時間的位置に対応して値が変化する関数を、前記新鮮度として用いて、前記所定のモデルの適応を行う

ことを特徴とする請求項 2 に記載のモデル適応装置。

【請求項 4】 前記パターン認識は、前記入力データの特徴空間における特徴分布に基づいて行われ、

前記モデル適応手段は、前記抽出データから得られる前記特徴分布を用いて、モデルの適応を行う

ことを特徴とする請求項 1 に記載のモデル適応装置。

【請求項 5】 前記モデル適応手段は、前記所定のモデルから前記抽出データが観測される、前記最尤法、複合統計量、または最短距離－最大分離定理に基づく所定の尺度が最大または最小になるように、モデルの適応を行う

ことを特徴とする請求項 1 に記載のモデル適応装置。

【請求項6】 前記モデル適応手段は、ニュートン下降(Newton Descent)法またはモンテカルロ(Monte Carlo)法により、前記最尤法、複合統計量、または最短距離-最大分離定理に基づく尺度を最大または最小にする前記所定のモデルのパラメータを求める

ことを特徴とする請求項5に記載のモデル適応装置。

【請求項7】 前記最短距離-最大分離定理に基づく尺度は、バタチャリア(Bhattacharyya)距離を用いて定義されるものである

ことを特徴とする請求項5に記載のモデル適応装置。

【請求項8】 前記入力データは、音声データである

ことを特徴とする請求項1に記載のモデル適応装置。

【請求項9】 前記所定のモデルは、音声区間でない区間における前記入力データを表す音響モデルである

ことを特徴とする請求項8に記載のモデル適応装置。

【請求項10】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識に用いられる前記モデルの適応を行うモデル適応方法であって、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出ステップと、

前記所定区間における抽出データを用い、最尤法、複合統計量、または最短距離-最大分離定理のうちのいずれかによって、前記所定のモデルの適応を行うモデル適応ステップと

を備えることを特徴とするモデル適応方法。

【請求項11】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識に用いられる前記モデルの適応を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出ステップと、

前記所定区間における抽出データを用い、最尤法、複合統計量、または最短距離-最大分離定理のうちのいずれかによって、前記所定のモデルの適応を行うモ

デル適応ステップと

を備えるプログラムが記録されている

ことを特徴とする記録媒体。

【請求項 1 2】 時系列の入力データを、所定数のモデルのうちのいずれかに分類するパターン認識装置であって、

前記入力データの特徴量を抽出する特徴抽出手段と、

前記所定数のモデルを記憶する記憶手段と、

前記入力データの特徴量を、前記所定数のモデルのうちのいずれかに分類する分類手段と、

所定のモデルに対応する、所定区間で観測される前記入力データを抽出し、抽出データとして出力するデータ抽出手段と、

前記所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、前記所定のモデルの適応を行うモデル適応手段と

を備えることを特徴とするパターン認識装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、モデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置に関し、特に、例えば、音声認識等を行う場合に用いて好適なモデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置に関する。

【0 0 0 2】

【従来の技術】

従来より、ノイズ環境下において発話された単語等を認識する方法が知られており、その代表的な方法としては、PMC(Parallel Model Combination)法、SS/NS S(Spectral Subtraction/Nonlinear Spectral Subtraction)法、SFE(Stochastic Feature Extraction)法等がある。

【0 0 0 3】

PMC法は、環境ノイズの情報を直接的に音響モデルに取り込んでいるので、認

識性能は良いが、その反面、計算コストが高い。即ち、PMC法では、（高度な演算を必要とするので、装置の規模が大型化したり、処理に要する時間が長くなったりする）。SS/NSS法では、音声データの特徴量を抽出する段階において、環境ノイズが除去される。したがって、SS/NSS法は、PMC法よりも計算コストが低く、現在、多く用いられている方法である。SFE法では、SS/NSS法と同様に、環境ノイズを含む音声信号の特徴量を抽出する段階において、環境ノイズが除去されるが、特徴量として、確率分布で表されるものが抽出される。SFE法では、このように、音声の特徴量が、特徴空間上の分布として抽出される点で、音声の特徴量が、特徴空間上の点として抽出されるSS/NSS法や、PMC法と異なる。

【 0 0 0 4 】

上述したいずれの方法においても、音声の特徴量の抽出後は、その特徴量が、予め登録されている複数の単語等に対応する音響モデルのうちのいずれに最も適合するかが判定され、最も適合する音響モデルに対応する単語が認識結果として出力される。

【 0 0 0 5 】

なお、SFE法については、本件出願人が先に出願した特開平11-133992号（特願平9-300979号）等に、その詳細が記載されている。また、PMC法、SS/NSS法、およびSFE法の性能比較等は、例えば、"H.Pao, H.Honda, K.Minamino, M.Omote, H.Ogawa and N.Iwahashi, Stochastic Feature Extraction for Improving Noise Robustness in Speech Recognition, Proceedings of the 8th Sony Research Forum, SRF98-234, pp.9-14, October 1998", "N.Iwahashi, H.Pao, H.Honda, K.Minamino and M.Omote, Stochastic Features for Noise Robust in Speech Recognition, ICASSP'98 Proceedings, pp.633-636, May, 1998", "N.Iwahashi, H.Pao(presented), H.Honda, K.Minamino and M.Omote, Noise Robust Speech Recognition Using Stochastic Representation of Features, ASJ'98-Spring Proceedings, pp.91-92, March, 1998", "N.iwahashi, H.Pao H.Honda, K.Minamino and M.Omote, Stochastic Represetation of Feature for Noise Robust Speech Recognition, Technical Report of IEICE, pp.19-24, SP97-97(1998-01)等に、その詳細が記載されている。

【 0 0 0 6 】

【発明が解決しようとする課題】

ところで、上述のSFE法等においては、音声認識の段階で、環境ノイズが直接的に反映されていないため、すなわち、環境ノイズの情報が直接的に音響モデルに取り込まれていないため、認識性能が低下する課題があった。

【 0 0 0 7 】

また、環境ノイズの情報が直接的に音響モデルに取り込まれていないことに起因して、音声認識が開始された時点から発話が始まるまでの時間が長くなるにつれて認識性能が、さらに低下する課題があった。

【 0 0 0 8 】

本発明はこのような状況に鑑みてなされたものであり、環境ノイズの情報を用いて音響モデルを補正することにより、音声認識が開始された時から発話が始まるまでの時間が長くなるに伴って認識性能が低下することを抑止することができるようにするものである。

【 0 0 0 9 】

【課題を解決するための手段】

本発明のモデル適応装置は、所定のモデルに対応する、所定区間で観測される入力データを抽出し、抽出データとして出力するデータ抽出手段と、所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、所定のモデルの適応を行うモデル適応手段とを備えることを特徴とする。

【 0 0 1 0 】

本発明のモデル適応方法は、所定のモデルに対応する、所定区間で観測される入力データを抽出し、抽出データとして出力するデータ抽出ステップと、所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、所定のモデルの適応を行うモデル適応ステップとを備えることを特徴とする。

【 0 0 1 1 】

本発明の記録媒体は、所定のモデルに対応する、所定区間で観測される入力デ

ータを抽出し、抽出データとして出力するデータ抽出ステップと、所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、所定のモデルの適応を行うモデル適応ステップとを備えるプログラムが記録されていることを特徴とする。

【0012】

本発明のパターン認識装置は、所定のモデルに対応する、所定区間で観測される入力データを抽出し、抽出データとして出力するデータ抽出手段と、所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、所定のモデルの適応を行うモデル適応手段とを備えることを特徴とする。

【0013】

本発明のモデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置においては、所定のモデルに対応する、所定区間で観測される入力データが抽出され、抽出データとして出力される。そして、所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、所定のモデルの適応が行われる。

【0014】

【発明の実施の形態】

図1は、本発明を適用した音声認識装置の一実施の形態の構成例を示している。

【0015】

この音声認識装置において、マイクロフォン1は、認識対象である発話音声、環境ノイズとともに集音し、フレーム化部2に出力する。フレーム化部2は、マイクロフォン1から入力される音声データをデジタル化する。さらに、フレーム化部2は、そのデジタル音声データを、所定の時間間隔（例えば、10ms）で取り出し、その取り出したデータを、1フレームのデータとして出力する。フレーム化部2が出力する1フレーム単位の音声データは、そのフレームを構成する時系列の音声データそれぞれをコンポーネントとする観測ベクトルaとして、ノイズ観測区間抽出部3、および特徴抽出部5に供給される。

【0016】

ここで、以下、適宜、第 t フレームの音声データである観測ベクトルを、 $a(t)$ と表す。

【0017】

ノイズ観測区間抽出部3は、フレーム化部2から入力されるフレーム単位の音声データを所定の時間（ $2N$ フレーム分以上）だけバッファリングし、図2に示すように、発話スイッチ4がオンとされたタイミング t_b から $2N$ フレーム分だけ以前のタイミング t_0 までをノイズ観測区間として、そのノイズ観測区間における $2N$ フレーム分の観測ベクトル a を抽出して、特徴抽出部5、および無音音響モデル補正部7に出力する。なお、本実施の形態では、ノイズ観測区間は、後述する特徴分布を抽出するためのノイズ観測区間 T_m と、音響モデルの適応を行うためのノイズ観測区間 T_n の2つに分けられており、ノイズ観測区間 T_m および T_n のいずれも、 N フレームとされている。但し、ノイズ観測区間 T_m と T_n のフレーム数は、同一である必要はない。

【0018】

発話スイッチ4は、ユーザが発話を開始するときにユーザによってオンとされ、発話を終了するときにオフとされる。したがって、図2に示すように、発話スイッチ4がオンとされたタイミング t_b 以前（ノイズ観測区間）の音声データには、発話音声は含まれず、環境ノイズだけが存在する。また、発話スイッチ4がオンとされたタイミング t_b から発話スイッチ4がオフとされるタイミング t_d までは、音声認識区間とされて、その音声認識区間の音声データが音声認識の対象とされる。

【0019】

特徴抽出部5は、ノイズ観測区間抽出部3から入力されるノイズ観測区間 T_m と T_n のうちの前半のノイズ観測区間 T_m の環境ノイズだけが存在する音声データに基づいて、フレーム化部2から入力される、タイミング t_b 以降の音声認識区間の観測ベクトル a から環境ノイズ成分を除去して、その特徴量を抽出する。

【0020】

すなわち、特徴抽出部5は、例えば、観測ベクトル a としての真（環境ノイズ

が除去された)の音声データをフーリエ変換し、そのパワースペクトラムを求め、そのパワースペクトラムの各周波数成分をコンポーネントとする特徴ベクトル y を算出する。なお、パワースペクトラムの算出方法は、フーリエ変換によるものに限定されるものではない。すなわち、パワースペクトラムは、その他、例えば、いわゆるフィルタバンク法などによって求めることも可能である。

【0021】

そして、特徴抽出部5は、観測ベクトル a としての音声データに含まれる真の音声を、その特徴量の空間(特徴ベクトル空間)に写像したときに得られる、その特徴ベクトル空間上の分布を表すパラメータ(以下、特徴分布パラメータと記述する) Z を、特徴ベクトル y と、ノイズ観測区間 T_m の環境ノイズに基づいて算出し、音声認識部6に供給する。

【0022】

図3は、図1の特徴抽出部5の詳細な構成例を示している。フレーム化部2から入力される観測ベクトル a は、特徴抽出部5において、パワースペクトラム分析部11に供給される。パワースペクトラム分析部11では、観測ベクトル a が、例えば、FFT(高速フーリエ変換)アルゴリズムによってフーリエ変換され、これにより、音声のパワースペクトラムが、特徴ベクトルとして抽出される。なお、ここでは、1フレームの音声データとしての観測ベクトル a が、 M 個のコンポーネントからなる特徴ベクトル(M 次元の特徴ベクトル)に変換されるものとする。

【0023】

ここで、第 t フレームの観測ベクトル $a(t)$ から得られる特徴ベクトルを $y(t)$ と表す。また、特徴ベクトル $y(t)$ のうち、真の音声のスペクトル成分を $x(t)$ と、環境ノイズのスペクトル成分を $u(t)$ と表す。この場合、真の音声のスペクトル成分 $x(t)$ は、次式(1)で表される。

【0024】

【数1】

$$x(t) = y(t) - u(t) \quad \dots (1)$$

ただし、ここでは、環境ノイズが不規則な特性を有し、また、観測ベクトル $a(t)$ としての音声データは、真の音声成分に環境ノイズを加算したものであると仮定している。

【0025】

一方、ノイズ観測区間抽出部 3 から入力されるノイズ観測区間 T_m における音声データとしての環境ノイズは、特徴抽出部 5 において、ノイズ特性算出部 13 に供給される。ノイズ特性算出部 13 では、ノイズ観測区間 T_m における環境ノイズの特性が求められる。

【0026】

すなわち、ここでは、音声認識区間における環境ノイズのパワースペクトラム $u(t)$ の分布が、その音声認識区間の直前のノイズ観測区間 T_m における環境ノイズと同一であり（または近似しており）、かつ、その分布が正規分布であると仮定して、ノイズ特性算出部 13 において、その正規分布を規定する、環境ノイズの平均値（平均ベクトル）と分散（分散マトリクス）（共分散行列）が求められる。

【0027】

なお、平均ベクトル μ' と分散マトリクス Σ' は、次式（2）にしたがって求めることができる。

【0028】

【数 2】

$$\begin{aligned}\mu'(i) &= \frac{1}{N} \sum_{t=1}^N y(t)(i) \\ \Sigma'(i, j) &= \frac{1}{N} \sum_{t=1}^N (y(t)(i) - \mu'(i))(y(t)(j) - \mu'(j)) \\ &\dots (2)\end{aligned}$$

ただし、 $\mu'(i)$ は、平均ベクトル μ' の i 番目のコンポーネントを表す（ $i = 1, 2, \dots, M$ ）。また、 $y(t)(i)$ は、第 t フレームの特徴ベクトルの i 番目のコンポーネントを表す。さらに、 $\Sigma'(i, j)$ は、分散マトリク

Σ' の、第 i 行、第 j 列のコンポーネントを表す ($j = 1, 2, \dots, M$)

。

【 0 0 2 9 】

ここで、計算量の低減のために、環境ノイズについては、特徴ベクトル y の各コンポーネントが、互いに無相関であると仮定する。この場合、次式に示すように、分散マトリクス Σ' は、対角成分以外は 0 となる。

【 0 0 3 0 】

【数 3】

$$\Sigma'(i, j) = 0, i \neq j \quad \dots (3)$$

【 0 0 3 1 】

ノイズ特性算出部 1 3 では、以上のようにして、環境ノイズの特性としての、正規分布を規定する平均ベクトル μ' および平均値 Σ' が求められ、特徴分布パラメータ算出部 1 2 に供給される。

【 0 0 3 2 】

一方、パワースペクトラム分析部 1 1 の出力、すなわち、環境ノイズを含む発話音声の特徴ベクトル y は、特徴分布パラメータ算出部 1 2 に供給される。特徴分布パラメータ算出部 1 2 では、パワースペクトラム分析部 1 1 からの特徴ベクトル y 、およびノイズ特性算出部 1 3 からの環境ノイズの特性に基づいて、真の音声のパワースペクトラムの分布（推定値の分布）を表す特徴分布パラメータが算出される。

【 0 0 3 3 】

すなわち、特徴分布パラメータ算出部 1 2 では、真の音声のパワースペクトラムの分布が正規分布であるとして、その平均ベクトル μ と分散マトリクス Ψ が、特徴分布パラメータとして、次式 (4) 乃至 (7) にしたがって計算される。

【 0 0 3 4 】

【数 4】

$$\begin{aligned}
 \xi(t)(i) &= E[x(t)(i)] \\
 &= E[y(t)(i) - u(t)(i)] \\
 &= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i)) \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\
 &= \frac{y(t)(i) \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) - \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\
 &= y(t)(i) - \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \quad \dots (4)
 \end{aligned}$$

【0 0 3 5】

【数 5】

 $i=j$ のとき

$$\begin{aligned}
 \Psi(t)(i, j) &= V[x(t)(i)] \\
 &= E[(x(t)(i))^2] - (E[x(t)(i)])^2 \\
 &= E[(x(t)(i))^2] - (\xi(t)(i))^2
 \end{aligned}$$

 $i \neq j$ のとき

$$\Psi(t)(i, j) = 0 \quad \dots (5)$$

【0 0 3 6】

【数 6】

$$E[(x(t)(i))^2] = E[(y(t)(i) - u(t)(i))^2]$$

$$\begin{aligned}
 &= \int_0^{y(t)(i)} (y(t)(i) - u(t)(i))^2 \frac{P(u(t)(i))}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} du(t)(i) \\
 &= \frac{1}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \times \left\{ (y(t)(i))^2 \int_0^{y(t)(i)} P(u(t)(i)) du(t)(i) \right. \\
 &\quad - 2y(t)(i) \int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i) \\
 &\quad \left. + \int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i) \right\} \\
 &= (y(t)(i))^2 - 2y(t)(i) \frac{\int_0^{y(t)(i)} u(t)(i) P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \\
 &\quad + \frac{\int_0^{y(t)(i)} (u(t)(i))^2 P(u(t)(i)) du(t)(i)}{\int_0^{y(t)(i)} P(u(t)(i)) du(t)(i)} \dots (6)
 \end{aligned}$$

【0037】

【数 7】

$$P(u(t)(i)) = \frac{1}{\sqrt{2\pi \Sigma'(i, i)}} e^{-\frac{1}{2\Sigma'(i, i)} (u(t)(i) - \mu'(i))^2} \dots (7)$$

【0038】

ここで、 $\xi(t)(i)$ は、第 t フレームにおける平均ベクトル $\xi(t)$ の i

番目のコンポーネントを表す。また、 $E[\]$ は、 $[\]$ 内の平均値を意味する。 $x(t)(i)$ は、第 t フレームにおける真の音声のパワースペクトラム $x(t)$ の i 番目のコンポーネントを表す。さらに、 $u(t)(i)$ は、第 t フレームにおける環境ノイズのパワースペクトラムの i 番目のコンポーネントを表し、 $P(u(t)(i))$ は、第 t フレームにおける環境ノイズのパワースペクトラムの i 番目のコンポーネントが $u(t)(i)$ である確率を表す。ここでは、環境ノイズの分布として正規分布を仮定しているので、 $P(u(t)(i))$ は、式 (7) に示したように表される。

【0039】

また、 $\Psi(t)(i, j)$ は、第 t フレームにおける分散 $\Psi(t)$ の、第 i 行、第 j 列のコンポーネントを表す。さらに、 $V[\]$ は、 $[\]$ 内の分散を表す。

【0040】

特徴分布パラメータ算出部 12 では、以上のようにして、各フレームごとに、平均ベクトル ξ および分散マトリクス Ψ が、真の音声の特徴ベクトル空間上での分布（ここでは、真の音声の特徴ベクトル空間上での分布が正規分布であると仮定した場合の、その分布）を表す特徴分布パラメータとして求められる。

【0041】

その後、音声認識区間の各フレームにおいて求められた特徴分布パラメータは、音声認識部 6 に出力される。すなわち、いま、音声認識区間が T フレームであったとし、その T フレームそれぞれにおいて求められた特徴分布パラメータを、 $z(t) = \{\xi(t), \Psi(t)\}$ ($t = 1, 2, \dots, T$) と表すと、特徴分布パラメータ算出部 12 は、特徴分布パラメータ（系列） $Z = \{z(1), z(2), \dots, z(T)\}$ を、音声認識部 6 に供給する。

【0042】

図 1 に戻り、音声認識部 6 は、特徴抽出部 5 から入力される特徴分布パラメータ Z を、所定数 K の音響モデルと 1 個の無音音響モデル（無音状態、即ち、環境ノイズのみが存在する状態を表す音響モデル）のうちのいずれかに分類し、その分類結果を、入力された音声の認識結果として出力する。すなわち、音声認識部 6 は、例えば、無音区間に対応する識別関数（特徴パラメータ Z が無音音響モデ

ルに分類されるかを識別するための関数)と、所定数 K の単語それぞれに対応する識別関数(特徴パラメータ Z がいずれの音響モデルに分類されるかを識別するための関数)とを記憶しており、各音響モデルの識別関数の値を、特徴抽出部5からの特徴分布パラメータ Z を引数として計算する。そして、音声認識部6は、その関数値(いわゆるスコア)が最大である音響モデル(単語、または無音(ノイズ))を認識結果として出力する。

【0043】

即ち、図4は、図1の音声認識部6の詳細な構成例を示している。特徴抽出部5の特徴分布パラメータ算出部12から入力される特徴分布パラメータ Z は、識別関数演算部21-1乃至21- K 、および識別関数演算部21- s に供給される。識別関数演算部21- k ($k=1, 2, \dots, K$)は、 K 個の音響モデルのうちの k 番目に対応する単語を識別するための識別関数 $G_k(Z)$ を記憶しており、特徴抽出部5からの特徴分布パラメータ Z を引数として、識別関数 $G_k(Z)$ を演算する。識別関数演算部21- s は、無音音響モデルに対応する無音区間を識別するための識別関数 $G_s(Z)$ を記憶しており、特徴抽出部5からの特徴分布パラメータ Z を引数として、識別関数 $G_s(Z)$ を演算する。

【0044】

なお、音声認識部6では、例えば、HMM(Hidden Markov Model)法を用いて、クラスとしての単語または無音の識別(認識)が行われる。

【0045】

HMM法について、図5を参照して説明する。同図において、HMMは、 H 個の状態 q_1 乃至 q_H を有しており、状態の遷移は、自身への遷移と、右隣の状態への遷移のみが許されている。また、初期状態は、最も左の状態 q_1 とされ、最終状態は、最も右の状態 q_H とされており、最終状態 q_H からの状態遷移は禁止されている。このように、自身よりも左にある状態への遷移のないモデルは、left-to-rightモデルと呼ばれ、音声認識では、一般に、left-to-rightモデルが用いられる。

【0046】

いま、HMMの k クラスを識別するためのモデルを、 k クラスモデルというすると、 k クラスモデルは、例えば、最初に状態 q_h にいる確率(初期状態確率)

$\pi_k(q_h)$ 、ある時刻（フレーム） t において、状態 q_i にいて、次の時刻 $t+1$ において、状態 q_j に状態遷移する確率（遷移確率） $a_k(q_i, q_j)$ 、および状態 q_i から状態遷移が生じるときに、その状態 q_i が、特徴ベクトル O を出力する確率（出力確率） $b_k(q_i)(O)$ によって規定される（ $h=1, 2, \dots, H$ ）。

【 0 0 4 7 】

そして、ある特徴ベクトル系列 O_1, O_2, \dots が与えられた場合、例えば、そのような特徴ベクトル系列が観測される確率（観測確率）が最も高いモデルのクラスが、その特徴ベクトル系列の認識結果とされる。

【 0 0 4 8 】

ここでは、この観測確率が、識別関数 $G_k(Z)$ によって求められる。すなわち、識別関数 $G_k(Z)$ は、特徴分布パラメータ（系列） $Z = \{z_1, z_2, \dots, z_T\}$ に対する最適状態系列（最適な状態の遷移のしていき方）において、そのような特徴分布パラメータ（系列） $Z = \{z_1, z_2, \dots, z_T\}$ が観測される確率を求めるものとして、次式（8）で与えられる。

【数 8】

$$G_k(Z) = \max_{q_1, q_2, \dots, q_T} \pi_k(q_1) \cdot b'_k(q_1)(z_1) \cdot a_k(q_1, q_2) \cdot b'_k(q_2)(z_2) \\ \cdot \dots \cdot a_k(q_{T-1}, q_T) \cdot b'_k(q_T)(z_T) \\ \dots (8)$$

【 0 0 4 9 】

ここで、 $b'_k(q_i)(z_j)$ は、出力が z_j で表される分布であるときの出力確率を表す。状態遷移時に各特徴ベクトルを出力する確率である出力確率 $b_k(s)(O_t)$ には、ここでは、例えば、特徴ベクトル空間上のコンポーネントに相関がないものとして、正規分布関数が用いられている。この場合、入力が z_t で表される分布であるとき、出力確率 $b'_k(s)(z_t)$ は、平均ベクトル $\mu_k(s)$ と分散マトリクス $\Sigma_k(s)$ とによって規定される確率密度関数 $P_k^m(s)(x)$ 、および第 t フレームの特徴ベクトル（ここでは、パワースペクトラム

) x の分布を表す確率密度関数 $P^f(t)(x)$ を用いて、次式により求めることができる。

【0050】

【数9】

$$\begin{aligned} b_k'(s)(z_t) &= \int P^f(t)(x) P_k^m(s)(x) dx \\ &= \prod_{i=1}^M P(s)(i)(\xi(t)(i), \Psi(t)(i, i)) \\ &\quad k=1, 2, \dots, K: s=q_1, q_2, \dots, q_T: T=1, 2, \dots, T \\ &\quad \dots (9) \end{aligned}$$

ただし、式(9)における積分の積分区間は、 M 次元の特徴ベクトル空間(ここでは、パワースペクトラム空間)の全体である。

【0051】

また、式(9)において、 $P(s)(i)(\xi(t)(i), \Psi(t)(i, i))$ は、次式(10)で表される。

【0052】

【数10】

$$\begin{aligned} P(s)(i)(\xi(t)(i), \Psi(t)(i, i)) \\ &= \frac{1}{\sqrt{2\pi(\sum_k(s)(i, i) + \Psi(t)(i, i))}} e^{-\frac{(\mu_k(s)(i) - \xi(t)(i))^2}{2(\sum_k(s)(i, i) + \Psi(t)(i, i))}} \\ &\quad \dots (10) \end{aligned}$$

ただし、 $\mu_k(s)(i)$ は、平均ベクトル $\mu_k(s)$ の i 番目のコンポーネントを、 $\sum_k(s)(i, i)$ は、分散マトリクス $\sum_k(s)$ の、第 i 行第 i 列のコンポーネントを、それぞれ表す。そして、 k クラスモデルの出力確率は、これらによって規定される。

【0053】

なお、HMMは、上述したように、初期状態確率 $\pi_k(q_h)$ 、遷移確率 $a_k(q_i$

、 q_j ）、および出力確率 $b_k(q_i)$ （ O ）によって規定されるが、これらは、学習用の音声データから特徴ベクトルを算出し、その特徴ベクトルを用いて、予め求めておくものとする。

【0054】

ここで、HMMとして、図5に示したものをを用いる場合には、常に、最も左の状態 q_1 から遷移が始まるので、状態 q_1 に対応する初期状態確率だけが1とされ、他の状態に対応する初期状態確率はすべて0とされる。また、出力確率は、式（9）、（10）から明らかなように、 $\Psi(t)(i, i)$ を0とすると、特徴ベクトルの分散を考慮しない場合の連続HMMにおける出力確率に一致する。

【0055】

なお、HMMの学習方法としては、例えば、Baum-Welchの再推定法などが知られている。

【0056】

図4に戻り、識別関数演算部21-k（ $k=1, 2, \dots, K$ ）は、 k クラスモデルについて、あらかじめ学習により求められている初期状態確率 $\pi_k(q_h)$ 、遷移確率 $a_k(q_i, q_j)$ 、および出力確率 $b_k(q_i)$ （ O ）によって規定される式（8）の識別関数 $G_k(Z)$ を記憶しており、特徴抽出部2からの特徴分布パラメータ Z を引数として、識別関数 $G_k(Z)$ を演算し、その関数値（上述した観測確率） $G_k(Z)$ を、決定部22に出力する。識別関数演算部21-sは、無音音響モデル補正部7から供給される初期状態確率 $\pi_s(q_h)$ 、遷移確率 $a_s(q_i, q_j)$ 、および出力確率 $b_s(q_i)$ （ O ）によって規定される、式（8）の識別関数 $G_k(Z)$ と同様の識別関数 $G_s(Z)$ を記憶しており、特徴抽出部2からの特徴分布パラメータ Z を引数として、識別関数 $G_s(Z)$ を演算し、その関数値（上述した観測確率） $G_s(Z)$ を、決定部22に出力する。

【0057】

決定部22では、識別関数演算部21-1乃至21-k、および識別関数演算部21-sそれぞれからの関数値 $G_k(Z)$ （ここでは、関数値 $G_s(Z)$ を含むものとする）に対して、例えば、次式（11）に示す決定規則を用いて、特徴分布パラメータ Z 、すなわち、入力された音声に属するクラス（音響モデル）が識

別される。

【0058】

【数11】

$$C(Z) = C_k, \text{ if } G_k(Z) = \max_i \{G_i(Z)\} \quad \dots (11)$$

ただし、 $C(Z)$ は、特徴分布パラメータ Z が属するクラスを識別する識別操作（処理）を行う関数を表す。また、式（11）の第2式の右辺における \max は、それに続く関数値 $G_i(Z)$ （ただし、ここでは、 $i = s, 1, 2, \dots, K$ ）の最大値を表す。

【0059】

決定部22は、式（11）にしたがって、クラスを決定すると、それを、入力された音声の認識結果として出力する。

【0060】

図1に戻り、無音音響モデル補正部7は、ノイズ観測区間抽出部3から入力されるノイズ観測区間 T_m と T_n のうちの後半のノイズ観測区間 T_n における音声データとしての環境ノイズに基づいて、音声認識部6に記憶されている無音音響モデルの適応を行うための新たな識別関数 $G_s(Z)$ を生成し、この新たな識別関数 $G_s(Z)$ によって、音声認識部6に記憶されている無音音響モデルの適応を行う。

【0061】

具体的には、無音音響モデル補正部7では、図6に示すように、ノイズ観測区間抽出部3から供給されるノイズ観測区間 T_n の音声データ（環境ノイズ）の N 個のフレームの各フレームについて、特徴ベクトル y が観測され、さらに、特徴抽出部5における場合と同様にして、次式で示すような特徴分布が生成される。

【0062】

【数12】

$$\{F_1(y), F_2(y), \dots, F_N(y)\} \quad \dots (12)$$

なお、特徴分布 $\{F_i(y), i = 1, 2, \dots, N\}$ は、確率密度関数(Probabilistic Density Function)であり、以下、無音特徴分布PDFとも記述する。

【0063】

次に、無音音響モデル補正部 7 は、無音特徴分布PDFを、次式に従い、図 7 に示すように、無音音響モデルに対応する確率分布 $F_s(y)$ に写像する。

【0064】

【数 13】

$$F_s(y) = V(F_1(y), F_2(y), \dots, F_N(y)) \quad \dots (13)$$

ただし、 V は無音特徴分布PDF $\{F_i(y), i = 1, 2, \dots, M\}$ を無音音響モデル $F_s(X)$ に写像する補正関数(写像関数)である。

【0065】

そして、無音音響モデル補正部 7 は、 $F_s(y)$ によって、音声認識部 6 が記憶している無音音響モデルを更新することにより、その無音音響モデルの適応を行う。

【0066】

ここで、無音音響モデルを表す確率分布 $F_s(y)$ が、例えば、平均値が μ_s で、共分散行列が Σ_s の正規分布であると仮定し、また、各フレームの特徴ベクトル y を構成するコンポーネントが無相関であると仮定すれば、無音特徴分布PDF $\{F_i(y), i = 1, 2, \dots, N\}$ の共分散行列 Σ_i は対角線行列となる。ただし、この仮定の前提条件として、無音音響モデルの共分散行列も対角線行列であることが必要である。したがって、ノイズ観測区間 T_n における各フレームの特徴ベクトル y を構成するコンポーネントが無相関であれば、無音特徴分布PDF $\{F_i(y), i = 1, 2, \dots, N\}$ は、各コンポーネントに対応する平均と分散を持つ正規分布 $N(\mu_i, \Sigma_i)$ となる。なお、 μ_i は $F_i(y)$ の平均値であり、 Σ_i は $F_i(y)$ の共分散行列である。

【0067】

以上の仮定の下、無音音響モデル補正部 7 は、無音特徴分布PDFを用い、最尤法(最大尤度法)、複合統計量(混合統計量)、または最短距離-最大分離定理

(最短距離法)のうちのいずれかによって、無音音響モデル $F_s(y)$ の適応を行う。

【0068】

即ち、最尤法によって、無音音響モデルの適応を行う場合には、図8に示すように、無音特徴分布PDF $\{F_i(y), i=1, 2, \dots, N\}$ を包含するような正規分布 $N(\mu_s, \Sigma_s)$ が、無音音響モデルとして求められる。

【0069】

ここで、図9に示すように、ノイズ観測区間 T_n における第 t フレームから得られる M 次元の特徴ベクトル y を、 $(y_1(t), y_2(t), \dots, y_M(t))$ と表す。また、特徴ベクトル $(y_1(t), y_2(t), \dots, y_M(t))$ から得られる特徴分布を Y_t と表し、特徴分布 Y_t を表す正規分布を $N(\mu_t, \Sigma_t)$ と表す。

【0070】

最大尤度法においては、正規分布 $N(\mu_s, \Sigma_s)$ で表される無音音響モデル $F_s(y)$ から、ノイズ観測区間 T_n において観測される無音特徴分布 Y_1, Y_2, \dots, Y_N が観測される度合いを表す尺度 L が、例えば、次式で定義される。

【0071】

【数14】

$$L \triangleq \log \Pr(Y_1, Y_2, \dots, Y_t, \dots, Y_N | N(\mu_s, \Sigma_s)) \quad \dots (14)$$

なお、 \log は自然対数を表し、 $\Pr(Y_1, Y_2, \dots, Y_N | N(\mu_s, \Sigma_s))$ は、無音音響モデル $N(\mu_s, \Sigma_s) (= F_s(y))$ から、無音特徴分布の系列 Y_1, Y_2, \dots, Y_N が観測される確率を表す。

【0072】

いま、無音特徴分布 Y_1, Y_2, \dots, Y_N が、それぞれ独立であるとする、式(14)の尺度 L は、次式で表すことができる。

【0073】

【数 15】

$$\begin{aligned}
 L &= \log \prod_{t=1}^N \Pr(Y_t \mid N(\mu_s, \Sigma_s)) \\
 &= \sum_{t=1}^N \log \Pr(Y_t \mid N(\mu_s, \Sigma_s)) \quad \dots (15)
 \end{aligned}$$

【0074】

式(15) (または式(14)の) 尺度は、その値が大きい程、無音音響モデルから、無音特徴分布 Y_1, Y_2, \dots, Y_N が観測される可能性が高いことを表すから、無音音響モデルの更新(適応)は、式(15)で表される尺度 L を最大(極大)にする無音音響モデル $N(\mu_s, \Sigma_s)$ によって行えば良い。従って、式(15)の尺度 L を最大にする無音音響モデル $F_s(y)$ としての正規分布 $N(\mu_s, \Sigma_s)$ を規定する平均 μ_s と分散 Σ_s を求めれば良いが、そのような平均 μ_s と分散 Σ_s は、式(14)の尺度 L を、 μ_s と分散 Σ_s でそれぞれ偏微分した偏微分結果を 0 にするものとなるから、式(16)を満たす平均 μ_s と分散 Σ_s を求めれば良いことになる。

【0075】

【数 16】

$$\begin{cases} \frac{\partial L}{\partial \mu_s} = 0 \\ \frac{\partial L}{\partial \Sigma_s} = 0 \end{cases} \quad \dots (16)$$

【0076】

ここで、式(16)を満たす平均 μ_s と分散 Σ_s を、式(17)のように表すこととすると、式(13)における補正関数(写像関数) V は、式(17)の関数 V_μ および V_Σ で規定されるものとなる。

【0077】

【数 17】

$$\begin{cases} \mu_s = V_\mu(\mu_t, \Sigma_t) \Big|_{t=1}^N \\ \Sigma_s = V_\Sigma(\mu_t, \Sigma_t) \Big|_{t=1}^N \end{cases} \dots (17)$$

【0078】

式(16)を解くために、無音特徴分布 $N(\mu_t, \Sigma_t)$ ($= Y_t$) を規定する平均 μ_t と分散 (共分散行列) Σ_t を、式(18)のように表すこととする。

【0079】

【数 18】

$$\mu_t = \begin{bmatrix} \mu_1(t) \\ \mu_2(t) \\ \vdots \\ \mu_M(t) \end{bmatrix} \quad \Sigma_t = \begin{bmatrix} \sigma_{11}^2(t) & \sigma_{12}^2(t) & \dots & \sigma_{1M}^2(t) \\ \sigma_{21}^2(t) & \sigma_{22}^2(t) & \dots & \sigma_{2M}^2(t) \\ \dots & \dots & \dots & \dots \\ \sigma_{M1}^2(t) & \sigma_{M2}^2(t) & \dots & \sigma_{MM}^2(t) \end{bmatrix} \dots (18)$$

なお、 t は、1 から N までの整数値であり、 $\sigma_{ij}^2(t)$ は、 i 次元と j 次元の共分散を表す。

【0080】

上述したように、ここでは、各フレームの特徴ベクトルを構成するコンポーネントが無相関であると仮定しており、その結果、無音特徴分布 $N(\mu_t, \Sigma_t)$ の共分散行列 Σ_t は対角線行列となるから、式(18)の Σ_t のコンポーネントのうち、 i と j とが等しくないもの (共分散) は 0 となる。従って、共分散行列 Σ_t は、次式で表されることになる。

【0081】

【数 19】

$$\Sigma_t = \begin{bmatrix} \sigma_{11}^2(t) & & & 0 \\ & \sigma_{22}^2(t) & & \\ & & \dots & \\ 0 & & & \sigma_{MM}^2(t) \end{bmatrix} \dots (19)$$

【0082】

無音音響モデルN (μ_s , Σ_s) についても、同様に、その平均 μ_s と分散 (共分散行列) Σ_s を、式 (20) のように表すこととする。

【0083】

【数20】

$$\mu_s = \begin{bmatrix} \mu_1(s) \\ \mu_2(s) \\ \vdots \\ \mu_M(s) \end{bmatrix} \quad \Sigma_s = \begin{bmatrix} \sigma_{11}^2(s) & \sigma_{12}^2(s) & \cdots & \sigma_{1M}^2(s) \\ \sigma_{21}^2(s) & \sigma_{22}^2(s) & \cdots & \sigma_{2M}^2(s) \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{M1}^2(s) & \sigma_{M2}^2(s) & \cdots & \sigma_{MM}^2(s) \end{bmatrix} \quad \cdots (20)$$

【0084】

そして、ここでは、無音音響モデルN (μ_s , Σ_s) を規定する共分散行列 Σ_s も、上述したように、対角線行列となるものと仮定しているから、式 (20) の Σ_s のコンポーネントのうち、i と j とが等しくないもの (共分散) は0となり、共分散行列 Σ_s は、式 (19) における場合と同様の次式で表されることになる。

【0085】

【数21】

$$\Sigma_s = \begin{bmatrix} \sigma_{11}^2(s) & & & 0 \\ & \sigma_{22}^2(s) & & \\ & & \cdots & \\ 0 & & & \sigma_{MM}^2(s) \end{bmatrix} \quad \cdots (21)$$

【0086】

ここで、以下、適宜、式 (19) の共分散行列 Σ_t を表すコンポーネントのサフィックスの一部を省略し、無音特徴分布N (μ_t , Σ_t) を規定する平均 μ_t と共分散行列 Σ_t を、次式のように表すこととする。

【0087】

【数 2 2】

$$\left\{ \begin{array}{l} \mu_t = \begin{bmatrix} \mu_1(t) \\ \vdots \\ \mu_M(t) \end{bmatrix} \\ \Sigma_t = \begin{pmatrix} \sigma_1^2(t) & & 0 \\ & \ddots & \\ 0 & & \sigma_M^2(t) \end{pmatrix} \end{array} \right. \quad \dots (22)$$

但し、 $t = 1, 2, \dots, N$ である。

【0 0 8 8】

同様に、以下、適宜、式 (2 1) の共分散行列 Σ_s を表すコンポーネントのサフィックスの一部を省略し、無音音響モデル $N(\mu_s, \Sigma_s)$ を規定する平均 μ_s と共分散行列 Σ_s を、次式のように表すこととする。

【0 0 8 9】

【数 2 3】

$$\left\{ \begin{array}{l} \mu_s = \begin{bmatrix} \mu_1(s) \\ \vdots \\ \mu_M(s) \end{bmatrix} \\ \Sigma_s = \begin{pmatrix} \sigma_1^2(s) & & 0 \\ & \ddots & \\ 0 & & \sigma_M^2(s) \end{pmatrix} \end{array} \right. \quad \dots (23)$$

【0 0 9 0】

この場合、式 (1 5) における無音特徴分布 $Y_t (= N(\mu_t, \Sigma_t))$ を、平均 μ_t と共分散行列 Σ_t で規定される確率密度関数と考えるとともに、無音音響モデル $N(\mu_s, \Sigma_s)$ を、平均 μ_s と共分散行列 Σ_s で規定される確率密度関数と考えると、式 (1 5) の尺度 L は、次式のように計算することができる。

【0091】

【数24】

$$\begin{aligned}
L &= \sum_{t=1}^N \log \frac{1}{(2\pi)^{M/2} |\Sigma_t + \Sigma_s|^{1/2}} \cdot \exp \left[-\frac{1}{2} (\mu_t - \mu_s)^T (\Sigma_t + \Sigma_s)^{-1} (\mu_t - \mu_s) \right] \\
&= -\frac{1}{2} MN \log 2\pi - \frac{1}{2} \sum_t |\Sigma_t + \Sigma_s| - \frac{1}{2} (\mu_t - \mu_s)^T (\Sigma_t + \Sigma_s)^{-1} (\mu_t - \mu_s) \\
&= -\frac{1}{2} MN \log 2\pi - \frac{1}{2} \sum_{t=1}^N \sum_{k=1}^M \log (\sigma_k^2(t) + \sigma_k^2(s)) - \frac{1}{2} \sum_{t=1}^N \sum_{k=1}^M \frac{(\mu_k(t) - \mu_k(s))^2}{\sigma_k^2(t) + \sigma_k^2(s)} \\
&\quad \dots (24)
\end{aligned}$$

但し、式(24)において、 $(\Sigma_t + \Sigma_s)^{-1}$ は、次式で表される。

【数25】

$$(\Sigma_t + \Sigma_s)^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(t) + \sigma_1^2(s)} & & 0 \\ & \dots & \\ 0 & & \frac{1}{\sigma_M^2(t) + \sigma_M^2(s)} \end{pmatrix} \dots (25)$$

【0092】

式(24)で表される尺度Lを、式(16)に示したように、平均 μ_s と共分散行列 Σ_s でそれぞれ偏微分すると、次式のようになる。

【0093】

【数26】

$$\begin{cases} \frac{\partial L}{\partial \mu_k(s)} = -\sum_{t=1}^N \frac{\mu_k(t) - \mu_k(s)}{\sigma_k^2(t) + \sigma_k^2(s)} \\ \frac{\partial L}{\partial \sigma_k^2(s)} = -\frac{1}{2} \sum_{t=1}^N \frac{1}{\sigma_k^2(t) + \sigma_k^2(s)} + \frac{1}{2} \sum_{t=1}^N \frac{(\mu_k(t) - \mu_k(s))^2}{(\sigma_k^2(t) + \sigma_k^2(s))^2} \end{cases} \dots (26)$$

但し、 $k=1, 2, \dots, M$ である。

【0094】

式(26)から、式(16)を満たす平均 $\mu_s(\mu_1(s), \mu_2(s), \dots, \mu_M(s))$ と、分散 $\Sigma_s(\sigma_1^2(s), \sigma_2^2(s), \dots, \sigma_M^2(s))$ は、次式を解くことで求めることができる。

【0095】

【数27】

$$\begin{cases} \sum_{t=1}^N \frac{\mu_k(t) - \mu_k(s)}{\sigma_k^2(t) + \sigma_k^2(s)} = 0 \\ \sum_{t=1}^N \left(\frac{1}{\sigma_k^2(t) + \sigma_k^2(s)} - \frac{(\mu_k(t) - \mu_k(s))^2}{(\sigma_k^2(t) + \sigma_k^2(s))^2} \right) = 0 \end{cases} \dots (27)$$

【0096】

ここで、以下、適宜、式(27)における $\mu_k(t)$ 、 $\mu_k(s)$ 、 $\sigma_k^2(t)$ 、 $\sigma_k^2(s)$ のサフィックスを簡略化し、それぞれを、式(28)のように表すこととする。

【0097】

【数28】

$$\begin{aligned} \mu_t &= \mu_k(t) \\ \mu_s &= \mu_k(s) \\ v_t &= \sigma_k^2(t) \\ v_s &= \sigma_k^2(s) \end{aligned} \dots (28)$$

【0098】

この場合、式(27)は、式(29)のようになる。

【0099】

【数 29】

$$\left\{ \begin{array}{l} \sum_{t=1}^N \frac{\mu_t - \mu_s}{v_t + v_s} = 0 \\ \sum_{t=1}^N \frac{1}{v_t + v_s} - \sum_t \frac{(\mu_t - \mu_s)^2}{(v_t + v_s)^2} = 0 \end{array} \right. \quad \dots (29)$$

【0100】

式(29)を変形すると、式(30)が得られる。

【0101】

【数 30】

$$\left\{ \begin{array}{l} \mu_s = \frac{\sum_{t=1}^N \frac{\mu_t}{v_t + v_s}}{\sum_{t=1}^N \frac{1}{v_t + v_s}} \\ \sum_{t=1}^N \frac{1}{v_t + v_s} - \sum_{t=1}^N \frac{\left(\mu_t - \frac{\sum_{t=1}^N \frac{\mu_t}{v_t + v_s}}{\sum_{t=1}^N \frac{1}{v_t + v_s}} \right)^2}{(v_t + v_s)^2} = 0 \end{array} \right. \quad \dots (30)$$

【0102】

式(30)において、 μ_s を求めるには、 v_s を求める必要があるが、 v_s を求めるにあたっては、例えば、ニュートン下降(Newton Descent)法や、モンテカルロ(Monte Carlo)法等を用いることができる。

【0103】

以上のような最尤法によって、無音音響モデルの適応を行う場合には、無音音響モデル補正部7では、例えば、図10のフローチャートにしたがった処理(モデルの適応処理)が行われる。

【0104】

即ち、ステップS1では、ノイズ観測区間 T_n の音声データ（ノイズ）から、無音特徴分布 $F_t(y)$ （= $N(\mu_t, \Sigma_t)$ ）が求められ、ステップS2に進み、その無音特徴分布の系列を観測する式（15）の尺度 L を最大にする、式（30）の分散 v_s が、ニュートン下降法またはモンテカルロ法によって求められる。さらに、ステップS3において、ステップS2で求められた分散 v_s を用い、式（30）にしたがって、平均 μ_s が求められる。そして、ステップS4に進み、ステップS3で求められた平均 μ_s と、ステップS2で求められた分散 v_s とで規定される正規分布 $N(\mu_s, v_s)$ に対応する識別関数 $G_s(Z)$ が生成され、音声認識部6（図4）の識別関数演算部21-sの識別関数が更新されて、処理を終了する。

【0105】

次に、複合統計量によって、無音音響モデルの適応を行う場合には、図11に示すように、複数の統計量としての無音特徴分布の集合 $\{F_i(y), i=1, 2, \dots, N\}$ を複合し、その結果得られる複合統計量の分布としての正規分布 $N(\mu_s, \Sigma_s)$ によって、無音音響モデル $F_s(y)$ が更新される。

【0106】

複合統計量を用いる場合においては、正規分布 $N(\mu_s, \Sigma_s)$ で表される無音音響モデル $F_s(y)$ から、ノイズ観測区間 T_n において観測される無音特徴分布 $F_1(y), F_2(y), \dots, F_N(y)$ が観測される度合いを表す尺度 L が、例えば、次式で定義される。

【0107】

【数 3 1】

$$\begin{aligned}
L &= \log \prod_i^N E(F_s(F_i(y))) \\
&= \sum_{i=1}^N \log \int_{\Omega_i} F_s(y) \cdot F_i(y) dy \\
&= \sum_{i=1}^N \left[-\frac{(\mu_s - \mu_i)^2}{2(v_s + v_i)} - \frac{1}{2} \log(2\pi)^M (v_s + v_i) \right] \\
&= -\frac{1}{2} MN \log 2\pi - \frac{1}{2} \sum_{i=1}^N \log(v_s + v_i) - \frac{1}{2} \sum_{i=1}^N \frac{(\mu_s - \mu_i)^2}{v_s + v_i} \\
&\quad \dots (31)
\end{aligned}$$

なお、式(31)において、 $F_s(F_i(y))$ は複合統計量を表し、 $E()$ はカッコ内の期待値を表す。また、式(31)の第2行目における積分の積分範囲は、無音特徴分布 $F_i(y)$ を得るのに用いたM次元の特徴ベクトル y の特徴ベクトル空間(本実施の形態では、パワースペクトラム空間) Ω_i の全体である。さらに、式(31)の第2行目から第3行目の変形は、無音特徴分布 $F_i(y)$ ($=N(\mu_t, \Sigma_t)$)を、平均 μ_t と共分散行列 Σ_t で規定される確率密度関数と考えるとともに、無音音響モデル $F_s(X)$ ($=N(\mu_s, \Sigma_s)$)を、平均 μ_s と共分散行列 Σ_s で規定される確率密度関数と考えることによって行うことができる。

【0108】

無音音響モデルの更新(適応)は、式(31)で表される尺度 L を最大(極大)にする無音音響モデル $N(\mu_s, \Sigma_s)$ によって行えば良い。そして、そのような無音音響モデル $N(\mu_s, \Sigma_s)$ を規定する平均 μ_s と分散 v_s は、式(31)の尺度 L を、 μ_s と分散 v_s でそれぞれ偏微分した偏微分結果を0にするものとなるから、式(32)を満たす平均 μ_s と分散 $v_s (= \sigma_s^2)$ を求めれば良いことになる。

【0109】

【数 3 2】

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial \mu_s} = 0 \\ \frac{\partial L}{\partial v_s} = 0 \end{array} \right. \quad \dots (32)$$

【0 1 1 0】

式 (3 2) に、式 (3 1) の尺度 L を代入して整理すると、式 (3 3) が得られる。

【0 1 1 1】

【数 3 3】

$$\left\{ \begin{array}{l} \sum_{i=1}^N \frac{\mu_s - \mu_i}{v_s + v_i} = 0 \\ \sum_{i=1}^N \left\{ \frac{1}{v_s + v_i} - \frac{(\mu_s - \mu_i)^2}{(v_s + v_i)^2} \right\} = 0 \end{array} \right. \quad \dots (33)$$

【0 1 1 2】

そして、式 (3 3) は、式 (3 4) のように変形することができる。

【数 34】

$$\left\{ \begin{array}{l} \mu_s = \frac{\sum_{i=1}^N \frac{1}{v_s + v_i}}{\sum_{i=1}^N \frac{\mu_i}{v_s + v_i}} \\ v_s = \left\{ v_s : \sum_{i=1}^N \left(\frac{1}{v_s + v_i} - \frac{\left(\frac{\sum_{i=1}^N \frac{1}{v_s + v_i}}{\sum_{i=1}^N \frac{\mu_i}{v_s + v_i}} \right)^2}{(v_s + v_i)^2} \right) = 0 \right\} \end{array} \right. \dots (34)$$

【0113】

式(34)において、 μ_s を求めるには、 v_s を求める必要があるが、 v_s を求めるにあたっては、最尤法における場合と同様に、ニュートン下降法やモンテカルロ法等を用いることができる。

【0114】

以上のような複合統計量を用いることによって、無音音響モデルの適応を行う場合には、無音音響モデル補正部7では、例えば、図12のフローチャートにしたがった処理(モデルの適応処理)が行われる。

【0115】

即ち、ステップS11では、ノイズ観測区間 T_n の音声データ(ノイズ)から、無音特徴分布 $F_t(y) (= N(\mu_t, \Sigma_t))$ が求められ、ステップS12に進み、その無音特徴分布の系列を観測する式(31)の尺度 L を最大にする、式(34)の分散 v_s が、ニュートン下降法またはモンテカルロ法によって求められる。さらに、ステップS13において、ステップS12で求められた分散 v_s を用い、式(34)にしたがって、平均 μ_s が求められる。そして、ステップS14に進み、ステップS13で求められた平均 μ_s と、ステップS12で求められた分散 v_s とで規定される正規分布 $N(\mu_s, v_s)$ に対応する識別関数 $G_s(Z$

）が生成され、音声認識部 6（図 4）の識別関数演算部 2 1 - s の識別関数が更新されて、処理を終了する。

【0 1 1 6】

次に、最短距離－最大分離定理によって、無音音響モデルの適応を行う場合には、図 1 3 に示すように、正規分布である無音特徴分布 $F_1(y)$ ($=N(\mu_1, \Sigma_1)$)、 $F_2(y)$ ($=N(\mu_2, \Sigma_2)$)、 \dots 、 $F_N(y)$ ($=N(\mu_N, \Sigma_N)$) それぞれからの距離 d_1, d_2, \dots, d_N の総和を最小にする正規分布 $N(\mu_s, \Sigma_s)$ によって、無音音響モデル $F_s(y)$ が更新される。

【0 1 1 7】

ある正規分布 $N(\mu_i, \Sigma_i)$ と、他の正規分布 $N(\mu_j, \Sigma_j)$ との間の距離 d_{ij} としては、例えば、バタチャリア(Bhattacharyya)距離や、マハラノビ(Mahalanobi)距離等を用いることができる。

【0 1 1 8】

バタチャリア距離によれば、正規分布 $N(\mu_i, \Sigma_i)$ と $N(\mu_j, \Sigma_j)$ との間の距離 d_{ij} は、次式によって求めることができる。

【0 1 1 9】

【数 3 5】

$$d_{ij} = \|N(\mu_i, \Sigma_i) - N(\mu_j, \Sigma_j)\|$$

$$= \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \log \frac{|(\Sigma_i + \Sigma_j)/2|}{|\Sigma_i|^{1/2} \cdot |\Sigma_j|^{1/2}}$$

$\dots (35)$

【0 1 2 0】

また、マハラノビ距離によれば、正規分布 $N(\mu_i, \Sigma_i)$ と $N(\mu_j, \Sigma_j)$ との間の距離 d_{ij} は、次式によって求めることができる。

【0 1 2 1】

【数 3 6】

$$d_{ij} = \|N(\mu_i, \Sigma_i) - N(\mu_j, \Sigma_j)\|$$

$$= (\mu_i - \mu_j)^T \Sigma^{-1} (\mu_i - \mu_j) \quad (\Sigma_i = \Sigma_j) \quad \dots (36)$$

【0 1 2 2】

ところで、式 (3 6) のマハラノビ距離は、距離 d_{ij} を求める 2 つの分布の分散が同一という前提の下で求められるため、即ち、正規分布 $N(\mu_i, \Sigma_i)$ と $N(\mu_j, \Sigma_j)$ それぞれの共分散行列 Σ_i と Σ_j とが同一であるという前提 ($\Sigma_i = \Sigma_j = \Sigma$) の下で求められるため、マハラノビ距離を採用する場合には、無音特徴分布 $F_1(y)$ を表す正規分布 $N(\mu_i, \Sigma_i)$ が制約を受けることになる。

【0 1 2 3】

そこで、本実施の形態では、式 (3 5) のバタチャリア距離を採用することとする。

【0 1 2 4】

最短距離-最大分離定理を用いる場合においては、無音音響モデル $F_s(y)$ としての正規分布 $N(\mu_s, \Sigma_s)$ から、ノイズ観測区間 T_n において観測される無音特徴分布 $F_1(X)$, $F_2(X)$, \dots , $F_N(X)$ が観測される度合いを表す尺度 L が、例えば、次式で定義される。

【0 1 2 5】

【数 3 7】

$$L \triangleq \sum_{i=1}^N \|N(\mu_i, \Sigma_i) - N(\mu_s, \Sigma_s)\|$$

$$= \sum_{i=1}^N \left\{ \frac{1}{4} \cdot \frac{(\mu_i - \mu_s)^2}{\sigma_i^2 + \sigma_s^2} + \frac{1}{2} \log \frac{M}{2} \cdot \frac{\sigma_i^2 + \sigma_s^2}{\sigma_i \sigma_s} \right\}$$

$$= \frac{1}{4} \sum_{i=1}^N \left\{ \frac{(\mu_i - \mu_s)^2}{\sigma_i^2 + \sigma_s^2} + 2 \log \frac{\sigma_i^2 + \sigma_s^2}{\sigma_i \sigma_s} \right\} + \frac{N}{2} \log \frac{M}{2} \quad \dots (37)$$

【0 1 2 6】

式 (3 7) の尺度 L が最小になる場合が、無音特徴分布と、無音音響モデルと

しての正規分布との間の距離を最も短くするから、無音音響モデルの更新（適応）は、式（37）で表される尺度 L を最小（極小）にする無音音響モデル $N(\mu_s, \Sigma_s)$ によって行えば良い。そして、そのような無音音響モデル $N(\mu_s, \Sigma_s)$ を規定する平均 μ_s と分散 σ_s^2 は、式（37）の尺度 L を、 μ_s と分散 σ_s^2 でそれぞれ偏微分した偏微分結果を0にするものとなるから、式（38）を満たす平均 μ_s と分散 σ_s^2 を求めれば良いことになる。

【0127】

【数38】

$$\begin{cases} \frac{\partial L}{\partial \mu_s} = 0 \\ \frac{\partial L}{\partial \sigma_s^2} = 0 \end{cases} \dots (38)$$

【0128】

式（38）に、式（37）の尺度 L を代入して計算すると、式（39）が得られる。

【0129】

【数39】

$$\begin{cases} \sum_{i=1}^N \frac{\mu_i - \mu_s}{\sigma_i^2 + \sigma_s^2} = 0 \\ \sum_{i=1}^N \left\{ \frac{(\mu_i - \mu_s)^2}{(\sigma_i^2 + \sigma_s^2)^2} - \frac{\sigma_i^2 - \sigma_s^2}{\sigma_s^2 (\sigma_i^2 + \sigma_s^2)} \right\} = 0 \end{cases} \dots (39)$$

【0130】

そして、式（39）は、式（40）のように変形することができる。

【数 40】

$$\left\{ \begin{array}{l} \mu_s = \frac{\sum_{i=1}^N \frac{\mu_i}{\sigma_i^2 + \sigma_s^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2 + \sigma_s^2}} \\ \sigma_s^2 = \left\{ \sigma_s^2 : \sum_{i=1}^N \left[\frac{\left(\frac{\sum_{i=1}^N \frac{\mu_i}{\sigma_i^2 + \sigma_s^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2 + \sigma_s^2}} - \mu_i \right)^2}{(\sigma_i^2 + \sigma_s^2)^2} - \frac{\sigma_i^2 - \sigma_s^2}{\sigma_s^2 (\sigma_i^2 + \sigma_s^2)} \right] = 0 \right\} \end{array} \right. \quad \dots (40)$$

【0131】

式(40)において、平均 μ_s を求めるには、分散 σ_s^2 を求める必要があるが、分散 σ_s^2 を求めるにあたっては、上述の最尤法における場合と同様に、ニュートン下降法やモンテカルロ法等を用いることができる。

【0132】

以上のような最短距離—最大分離定理を用いることによって、無音音響モデルの適応を行う場合には、無音音響モデル補正部7では、例えば、図14のフローチャートにしたがった処理(モデルの適応処理)が行われる。

【0133】

即ち、ステップS21では、ノイズ観測区間 T_n の音声データ(ノイズ)から、無音特徴分布 $F_t(y) (=N(\mu_t, \Sigma_t))$ が求められ、ステップS22に進み、その無音特徴分布の系列を観測する式(37)の尺度 L を最小にする、式(40)の分散 σ_s^2 が、ニュートン下降法またはモンテカルロ法によって求められる。さらに、ステップS23において、ステップS22で求められた分散 σ_s^2 を用い、式(40)にしたがって、平均 μ_s が求められる。そして、ステップS24に進み、ステップS23で求められた平均 μ_s と、ステップS22で求められた分散 σ_s^2 とで規定される正規分布に対応する識別関数 $G_s(Z)$ が生成され、音声認識部6(図4)の識別関数演算部21-sの識別関数が更新されて、処

理を終了する。

【0134】

次に、図1の音声認識装置の動作について説明する。

【0135】

フレーム化部2には、マイクロフォン1で集音された音声データ（環境ノイズを含む認識対象の発話音声）が入力され、そこでは、音声データがフレーム化され、各フレームの音声データは、観測ベクトル a として、ノイズ観測区間抽出部3、および特徴抽出部5に順次供給される。ノイズ観測区間抽出部3では、発話スイッチ4がオンとされたタイミング t_b 以前のノイズ観測区間 T_m と T_n の音声データ（環境ノイズ）が抽出されて、特徴抽出部5および無音音響モデル補正部7に供給される。

【0136】

無音音響モデル補正部7では、ノイズ観測区間 T_m と T_n の音声データとしての環境ノイズに基づいて、上述した最尤法、複合統計量、または最短距離-最大分離定理のうちのいずれかを用いることによって、無音音響モデルの更新（適応）が行われ、音声認識部6に供給される。音声認識部6では、無音音響モデル補正部7から供給される無音音響モデルとしての識別関数によって、それまで記憶されていた無音音響モデルに対応する識別関数が更新される。即ち、無音音響モデルの適応が行われる。

【0137】

一方、特徴抽出部5では、フレーム化部2からの観測ベクトル a としての音声データが音響分析され、その特徴ベクトル y が求められる。さらに、特徴抽出部5では、求められた特徴ベクトル y と、ノイズ観測区間 T_m の音声データ（環境ノイズ）に基づいて、特徴ベクトル空間における分布を表す特徴分布パラメータ Z が算出され、音声認識部6に供給される。音声認識部6では、特徴抽出部5からの特徴分布パラメータを用いて、無音および所定数 K の単語それぞれに対応する音響モデルの識別関数の値が演算され、その関数値が最大となる音響モデルが、音声の認識結果として出力される。

【0138】

以上のように、観測ベクトル a としての音声データが、その特徴量の空間である特徴ベクトル空間における分布を表す特徴分布パラメータ Z に変換されるので、その特徴分布パラメータは、音声データに含まれるノイズの分布特性を考慮したものとなっており、また、無音を識別（認識）するための無音音響モデルに対応する識別関数が、発話直前のノイズ観測区間 T_n の音声データに基づいて更新されているので、音声認識率を大きく向上させることが可能となる。

【0139】

さらに、無音音響モデルの適応を行わない場合には、発話スイッチ 4 がオンとされてから発話が始まるまでの無音区間 T_s （図 2）が長くなるのに伴って、音声認識率が著しく低下するが、無音音響モデルの適応を行う場合には、無音区間 T_s が長くなっても、音声認識率の低下を僅かなものに抑えることができ、従って、無音区間 T_s の長短にほとんど影響を受けない認識性能を実現することができる。

【0140】

ところで、無音特徴分布 $F_i(y)$ ($=N(\mu_i, \sigma_i^2)$) を用いた、最尤法、複合統計量、または最短距離－最大分離定理による無音音響モデルの適応では、ノイズ観測区間 T_n （図 2）における N フレームそれぞれから得られる時系列の無音特徴分布 $F_1(y)$, $F_2(y)$, \dots , $F_N(y)$ を平等に取り扱っている。

【0141】

しかしながら、音声認識区間における環境ノイズは、厳密には、音声認識区間の直前のノイズ観測区間 T_n における環境ノイズと同一ではなく、さらに、一般には、ノイズ観測区間 T_n における環境ノイズは、音声認識区間（の開始時刻 t_c ）から離れるほど、音声認識区間における環境ノイズとは異なるものとなると推測される。

【0142】

従って、ノイズ観測区間 T_n （図 2）における N フレームそれぞれから得られる時系列の無音特徴分布 $F_1(y)$, $F_2(y)$, \dots , $F_N(y)$ は、平等に扱うのではなく、音声認識区間に近いものほど重みをおいて扱うべきであり（音

声認識区間から遠いものほど重みをおかずに扱うべきであり）、そのようにすることで、音声認識精度をより向上させる無音音響モデルの適応（補正または更新）が可能となる。

【0143】

そこで、ノイズ観測区間 T_n において得られる無音特徴分布 $F_1(y)$ ， $F_2(y)$ ， \dots ， $F_N(y)$ について、その新しさ（ここでは、音声認識区間への近さに相当する）を表す新鮮度を導入することとし、この新鮮度を考慮して、無音音響モデルの適応を行う方法について説明する。

【0144】

図15は、新鮮度を考慮して、無音音響モデルの適応を行う、図1の無音音響モデル補正部7の構成例を示している。

【0145】

新鮮度関数記憶部31は、上述したような新鮮度を表す関数である新鮮度関数（を規定するパラメータ）を記憶している。

【0146】

補正部32には、ノイズ観測区間抽出部3が出力する、ノイズ観測区間 T_n における音声データ（ノイズ）としての観測ベクトルの系列（ここでは、Nフレームの音声データ）が入力されるようになっており、補正部32は、この観測ベクトルから、無音特徴分布 $F_1(y)$ ， $F_2(y)$ ， \dots ， $F_N(y)$ を得て、これらと、新鮮度関数記憶部31に記憶されている新鮮度関数に基づいて、無音音響モデルの適応を行う。

【0147】

ここで、無音特徴分布 $F_1(y)$ ， $F_2(y)$ ， \dots ， $F_N(y)$ は、ノイズ観測区間 T_n におけるNフレームそれぞれで観測される離散値であり、無音音響モデル補正部7が、離散値を処理するシステムであれば、離散値である無音特徴分布 $F_1(y)$ ， $F_2(y)$ ， \dots ， $F_N(y)$ をそのまま用いることができる。しかしながら、無音音響モデル補正部7が、連続値を処理するシステムである場合には、例えば、図16に示すように、離散値である無音特徴分布 $F_1(y)$ ， $F_2(y)$ ， \dots ， $F_N(y)$ を、連続変換器で連続値に変換してから、無音

音響モデル補正部 7 で処理する必要がある。離散値を連続値に変換する方法としては、例えば、スプライン関数(Spline Function)によって近似を行う方法がある。

【0 1 4 8】

なお、離散値とは、ある有限の観測区間において、離散的な時刻で観測される有限個の観測値であり、連続値とは、ある有限（または無限）の観測区間の任意の時刻で観測される無限個の観測値であり、ある関数によって表現される。

【0 1 4 9】

無音音響モデルの適応に用いる無音特徴分布が離散値である場合には、新鮮度関数も離散値の関数となり、無音特徴分布が連続値である場合には、新鮮度関数も連続値の関数となる。

【0 1 5 0】

次に、新鮮度関数、およびそれを用いた無音音響モデルの適応について説明する。

【0 1 5 1】

新鮮度関数 $F(x)$ は、例えば、式 (4 1) 乃至 (4 3) に示すように定義することができる。

【0 1 5 2】

【数 4 1】

$$F(x) = 0 \quad \text{if } x \notin \Omega_{\text{obs}} \quad \dots (41)$$

【数 4 2】

$$F(x_2) \geq F(x_1) \quad \text{if } x_2 \geq x_1 \quad \dots (42)$$

【数 4 3】

$$\int_{\Omega_{\text{obs}}} F(x) dx \equiv 1 \quad \dots (43)$$

但し、 Ω_{obs} は、無音特徴分布の観測区間を表し、本実施の形態では、ノイズ観

測区間 T_n に相当する。

【0153】

式(41)により、新鮮度関数 $F(x)$ は、観測区間 Ω_{obs} 以外では0となる。また、式(42)により、新鮮度関数 $F(x)$ は、観測区間 Ω_{obs} において、時間の経過とともに増加するか、または変化しない関数であり、従って、新鮮度関数 $F(x)$ は、基本的に、音声認識区間(図2)に近づくほど、大きな値となる。さらに、式(43)により、新鮮度関数 $F(x)$ は、観測区間 Ω_{obs} に亘って積分した場合に、その積分値が1となる関数である。式(41)乃至(43)から、新鮮度関数 $F(x)$ は、例えば、図17に示すようになる。

【0154】

ここで、本実施の形態では、新鮮度関数 $F(x)$ は、後述する式(44)に示すように、無音特徴分布に乗算する乗数として用いられる。従って、新鮮度関数 $F(x)$ は、その値が正または負のときには、それが乗数として乗算される無音特徴分布に対する重みとして作用する。また、新鮮度関数 $F(x)$ は、その値が0のときは、それが乗数として乗算される無音特徴分布を無効とし、無音音響モデルの適応に影響を与えないように作用する。

【0155】

図15の補正部32では、以上のような新鮮度関数 $F(x)$ と、無音特徴分布 $F_1(y)$, $F_2(y)$, \dots , $F_N(y)$ とを用いて、基本的には、式(44)にしたがって、適応後の無音音響モデル $F_s(y)$ が求められる。

【0156】

【数44】

$$F_s(y) = V(F(1)F_1(y), F(2)F_2(y), \dots, F(N)F_N(y)) \quad \dots (44)$$

【0157】

式(44)によれば、無音特徴分布が、音声認識区間に近いものほど重みをおいて扱われて、無音音響モデルの適応が行われ、その結果、音声認識精度をより向上させることが可能となる。

【0158】

以上、本発明を適用した音声認識装置について説明したが、このような音声認

識装置は、例えば、音声入力可能なカーナビゲーション装置、その他各種の装置に適用可能である。

【 0 1 5 9 】

なお、本実施の形態では、ノイズの分布特性を考慮した特徴分布パラメータを求めるようにしたが、このノイズには、例えば、発話を行う環境下における外部からのノイズの他、例えば、電話回線その他の通信回線を介して送信されてくる音声の認識を行う場合には、その通信回線の特性なども含まれる。

【 0 1 6 0 】

また、本発明は、音声認識の他、画像認識その他のパターン認識を行う場合にも適用可能である。

【 0 1 6 1 】

さらに、本実施の形態では、特徴空間における分布として表される無音特徴分布を用いて、無音音響モデルの適応を行うようにしたが、無音音響モデルの適応は、特徴空間における点として表されるノイズの特徴量を用いて行うことも可能である。

【 0 1 6 2 】

また、本実施の形態では、ノイズを表す無音音響モデルの適応を行うようにしたが、上述した適応手法は、他の音響モデルの適用にも用いることが可能である。

【 0 1 6 3 】

次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【 0 1 6 4 】

そこで、図 1 8 は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【 0 1 6 5 】

プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク

ク 1 0 5 や ROM 1 0 3 に予め記録しておくことができる。

【 0 1 6 6 】

あるいはまた、プログラムは、フロッピーディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体 1 1 1 に、一時的あるいは永続的に格納（記録）しておくことができる。このようなリムーバブル記録媒体 1 1 1 は、いわゆるパッケージソフトウェアとして提供することができる。

【 0 1 6 7 】

なお、プログラムは、上述したようなリムーバブル記録媒体 1 1 1 からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部 1 0 8 で受信し、内蔵するハードディスク 1 0 5 にインストールすることができる。

【 0 1 6 8 】

コンピュータは、CPU(Central Processing Unit) 1 0 2 を内蔵している。CPU 1 0 2 には、バス 1 0 1 を介して、入出力インタフェース 1 1 0 が接続されており、CPU 1 0 2 は、入出力インタフェース 1 1 0 を介して、ユーザによって、キーボードやマウス等で構成される入力部 1 0 7 が操作されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory) 1 0 3 に格納されているプログラムを実行する。あるいは、また、CPU 1 0 2 は、ハードディスク 1 0 5 に格納されているプログラム、衛星若しくはネットワークから転送され、通信部 1 0 8 で受信されてハードディスク 1 0 5 にインストールされたプログラム、またはドライブ 1 0 9 に装着されたりリムーバブル記録媒体 1 1 1 から読み出されてハードディスク 1 0 5 にインストールされたプログラムを、RAM(Random Access Memory) 1 0 4 にロードして実行する。これにより、CPU 1 0 2 は、上述したブロック図の構成により行われる処理を行う。そして、CPU 1 0 2 は、その処理結

果を、必要に応じて、例えば、入出力インタフェース 1 1 0 を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部 1 0 6 から出力、あるいは、通信部 1 0 8 から送信、さらには、ハードディスク 1 0 5 に記録等させる。

【 0 1 6 9 】

ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【 0 1 7 0 】

また、プログラムは、1 のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【 0 1 7 1 】

【発明の効果】

本発明のモデル適応装置およびモデル適応方法、記録媒体、並びにパターン認識装置によれば、所定のモデルに対応する、所定区間で観測される入力データが抽出され、抽出データとして出力される。そして、所定区間における抽出データを用い、最尤法、複合統計量、または最短距離－最大分離定理のうちのいずれかによって、所定のモデルの適応が行われる。従って、そのモデルを用いてパターン認識を行うことで、認識性能を向上させることが可能となる。

【図面の簡単な説明】

【図 1】

本発明を適用した音声認識装置の構成例を示すブロック図である。

【図 2】

図 1 のノイズ観測区間抽出部 3 の動作を説明するための図である。

【図 3】

図 1 の特徴抽出部 5 の詳細な構成例を示すブロック図である。

【図 4】

図 1 の音声認識部 6 の詳細な構成例を示すブロック図である。

【図 5】

HMM を示す図である。

【図 6】

ノイズ観測区間 T_n から得られる特徴ベクトル y と特徴分布 $F_i(y)$ を示す図である。

【図 7】

無音特徴分布 PDF を、無音音響モデルに対応する確率分布 $F_s(y)$ に写像する様子を示す図である。

【図 8】

最尤法による無音音響モデルの適応の概要を示す図である。

【図 9】

ノイズ観測区間 T_n から得られる特徴ベクトルと、特徴分布 Y_t としての正規分布 $N(\mu_t, \Sigma_t)$ を示す図である。

【図 10】

最尤法による無音音響モデルの適応処理を説明するフローチャートである。

【図 11】

複合統計量を用いた無音音響モデルの適応の概要を示す図である。

【図 12】

複合統計量を用いた無音音響モデルの適応処理を説明するフローチャートである。

【図 13】

最短距離－最大分離定理を用いた無音音響モデルの適応の概要を示す図である。

【図 14】

最短距離－最大分離定理を用いた無音音響モデルの適応処理を説明するフローチャートである。

【図 15】

図 1 の無音音響モデル補正部 7 の構成例を示すブロック図である。

【図 1 6】

離散値を連続値に変換する様子を示す図である。

【図 1 7】

新鮮度関数 $F(x)$ を示す図である。

【図 1 8】

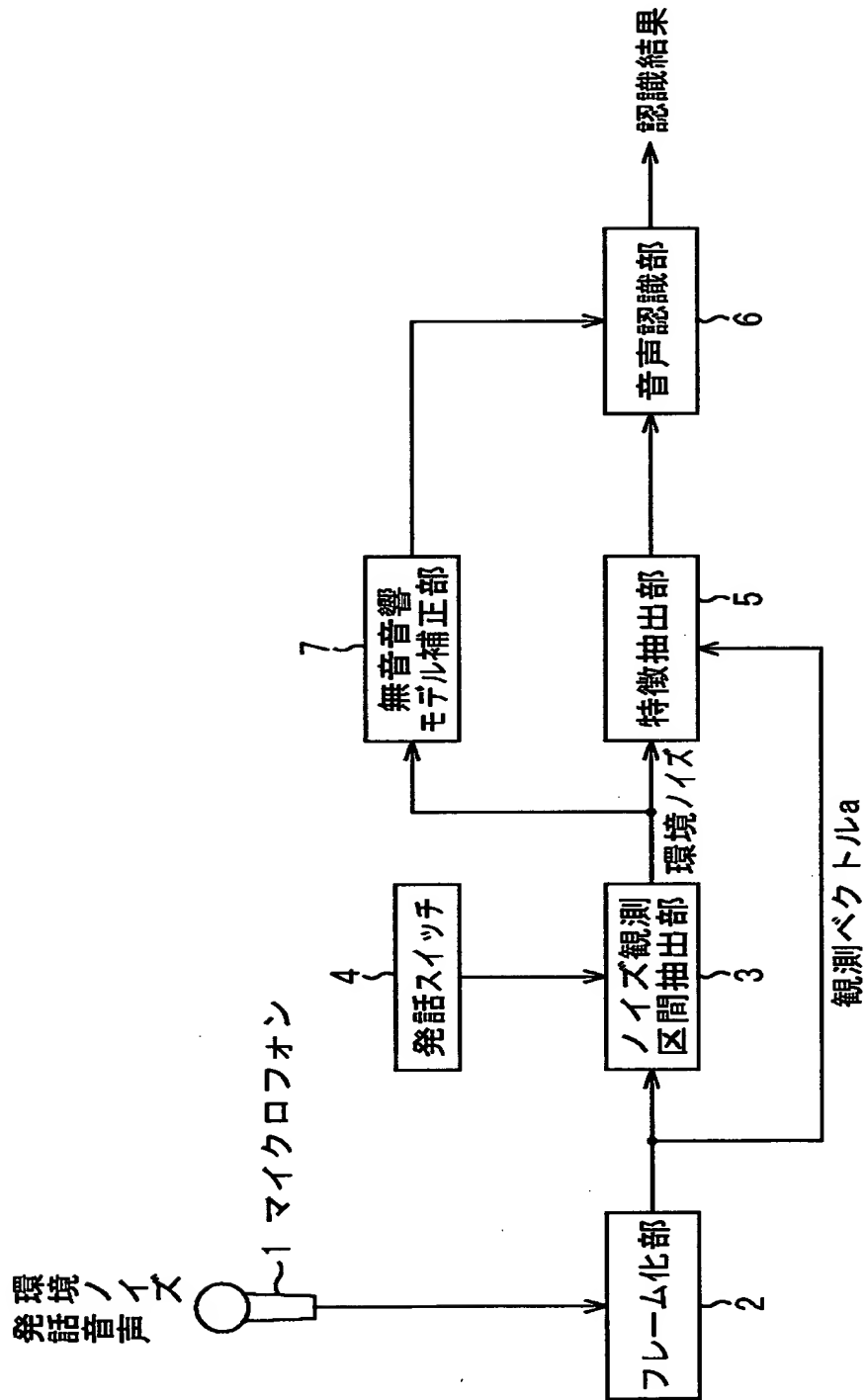
本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

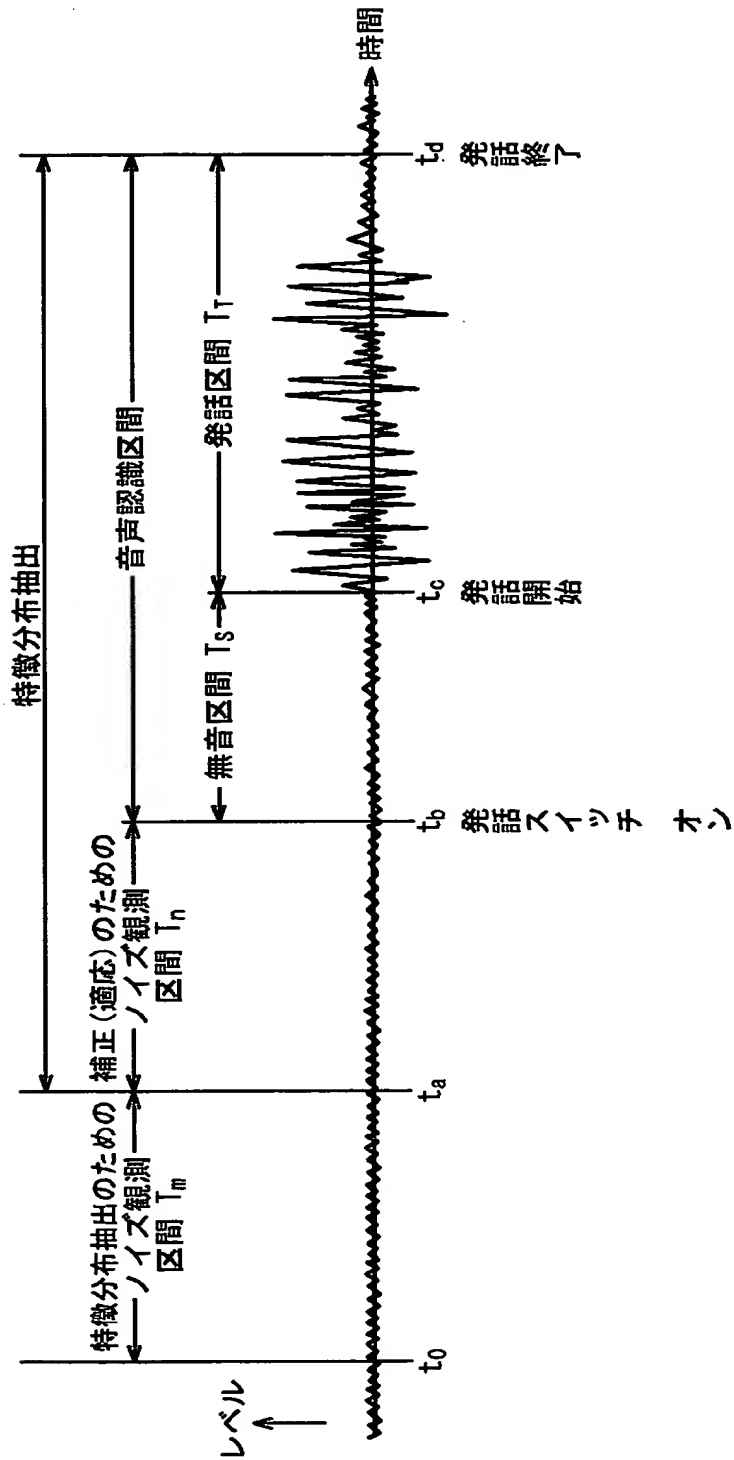
1 マイクロフォン, 2 フレーム化部, 3 ノイズ観測区間抽出部,
4 発話スイッチ, 5 特徴抽出部, 6 音声認識部, 7 無音音響モデル補正部,
31 新鮮度関数記憶部, 32 補正部, 101 バス, 102 CPU, 103 ROM, 104 RAM, 105 ハードディスク, 106 出力部,
107 入力部, 108 通信部, 109 ドライブ, 110 入出力インタフェース, 111 リムーバブル記録媒体

【書類名】 図面

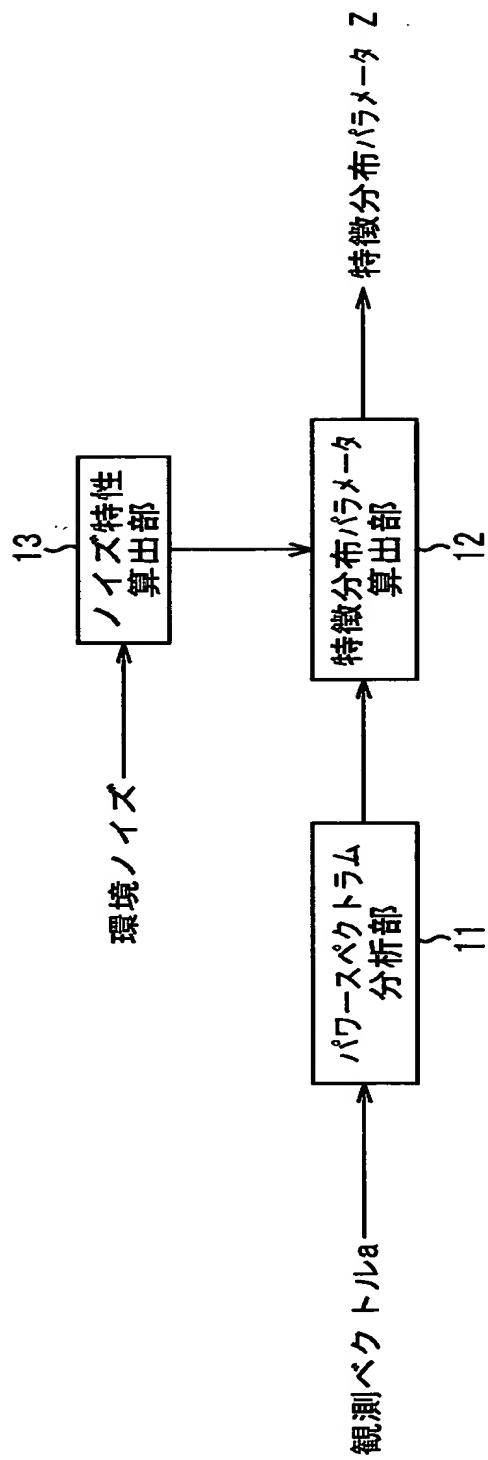
【図 1】



【図 2】

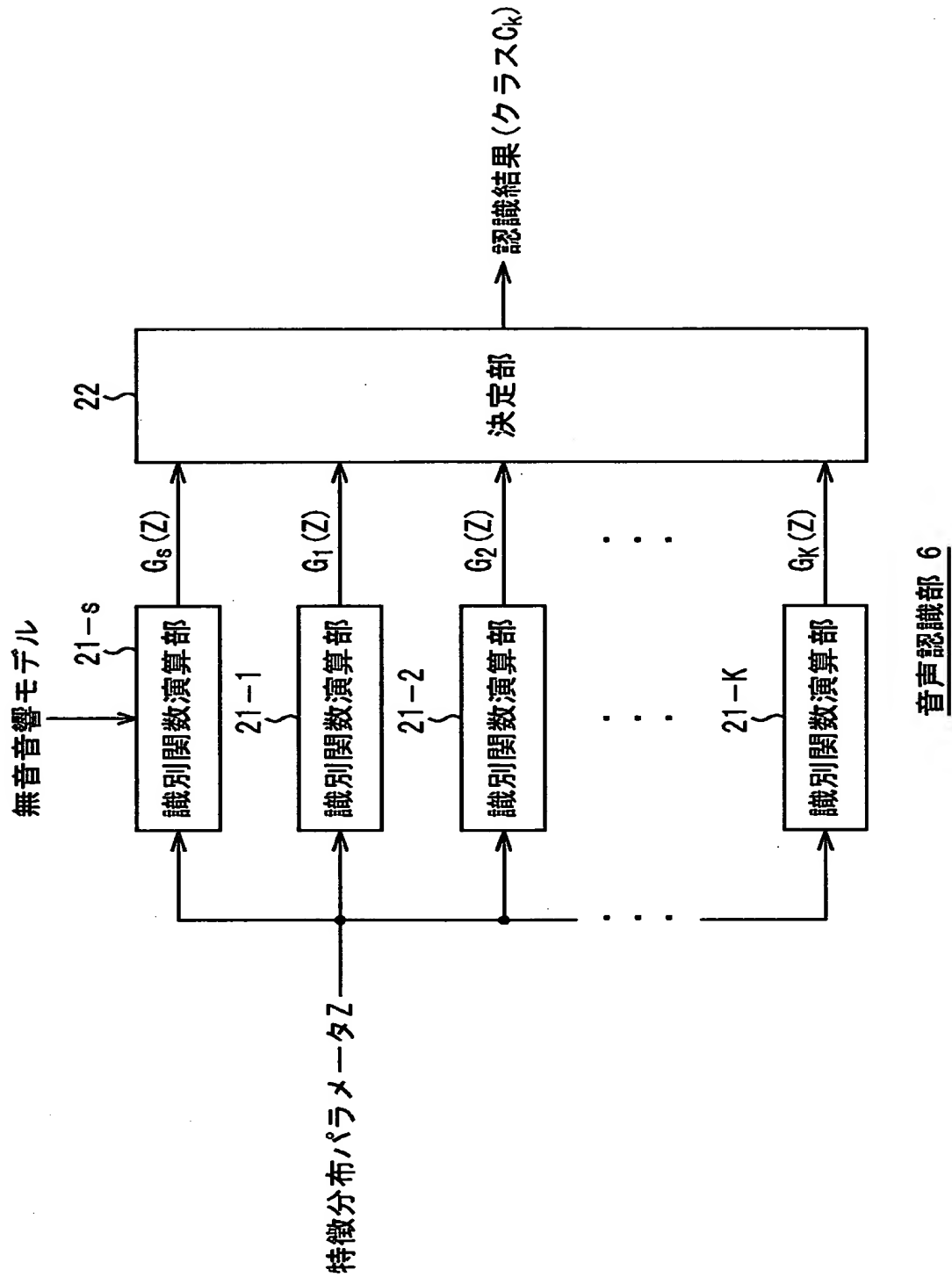


【図 3】

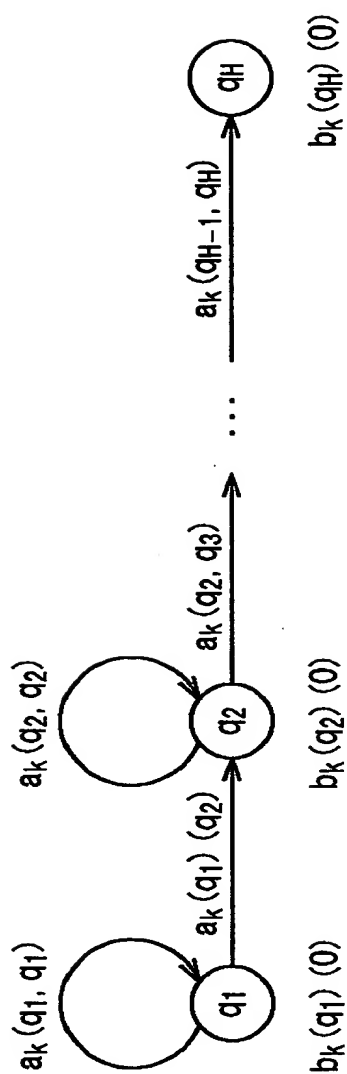


特徴抽出部 5

【図 4】

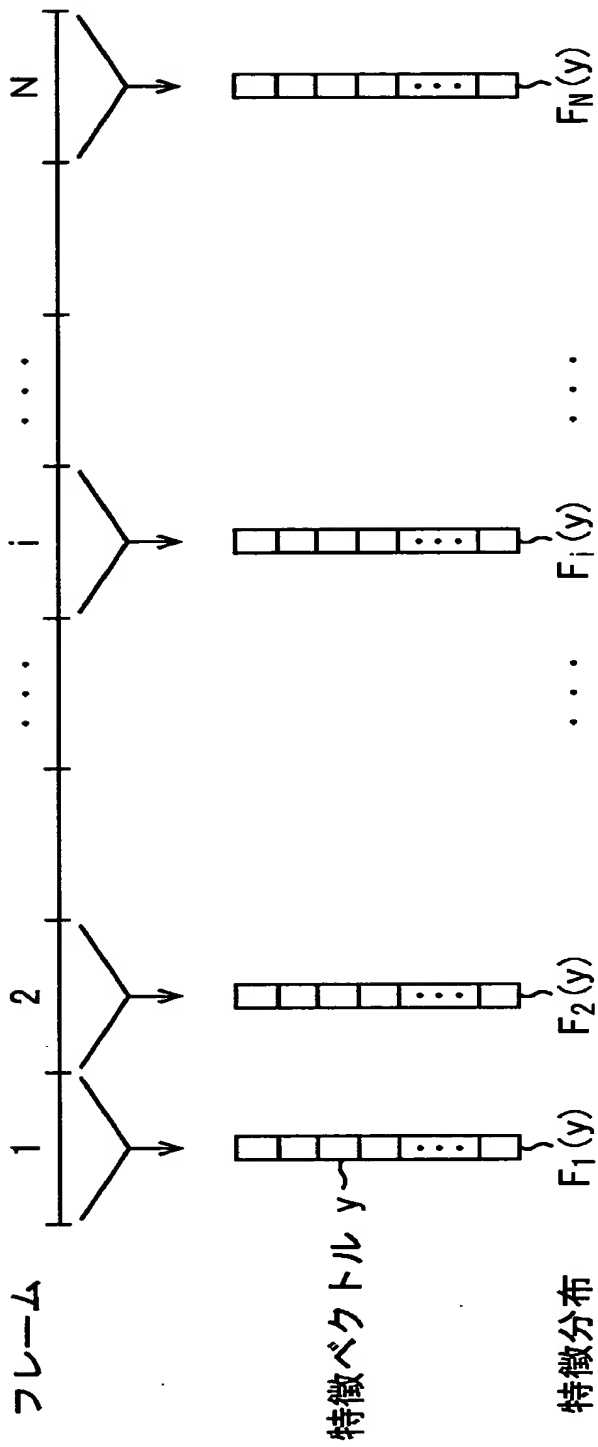


【図 5】

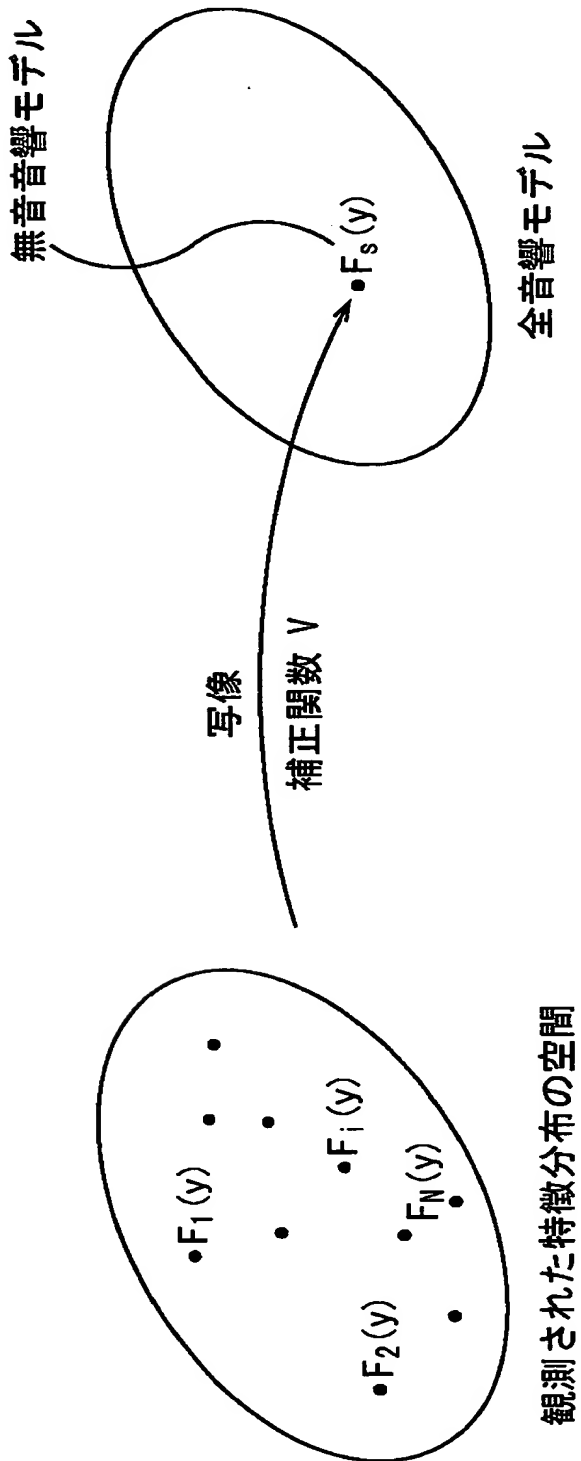


HMM (left-to-right モデル)

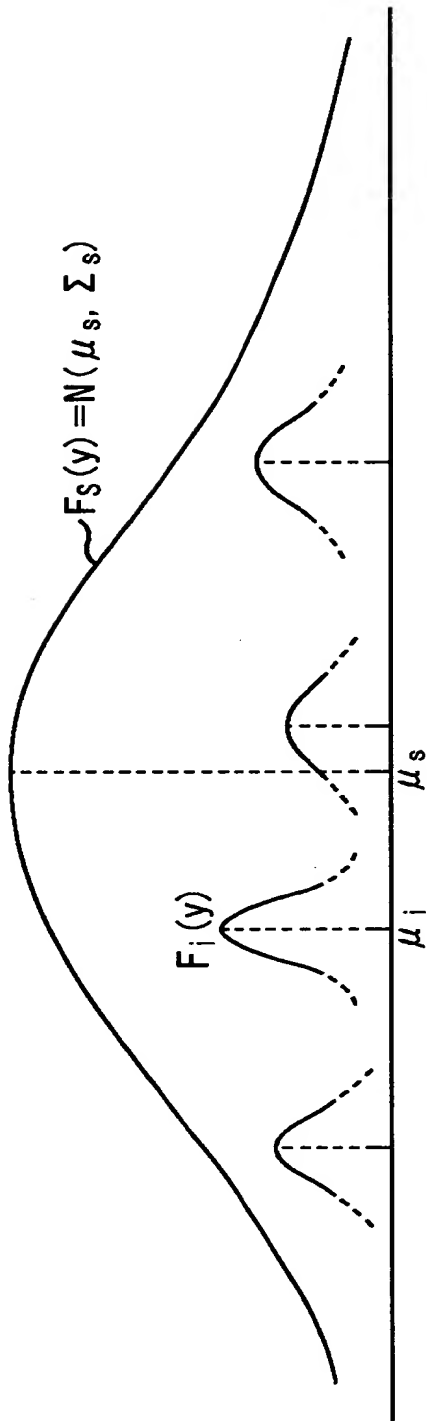
【図6】



【図 7】

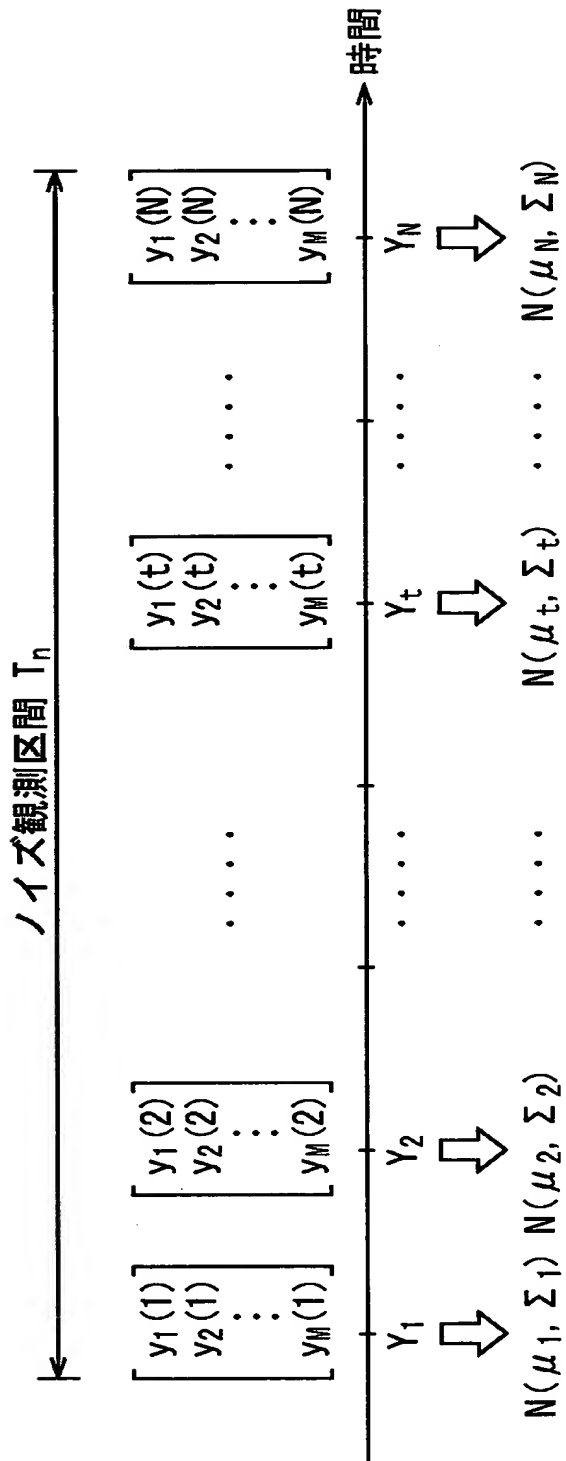


【図 8】

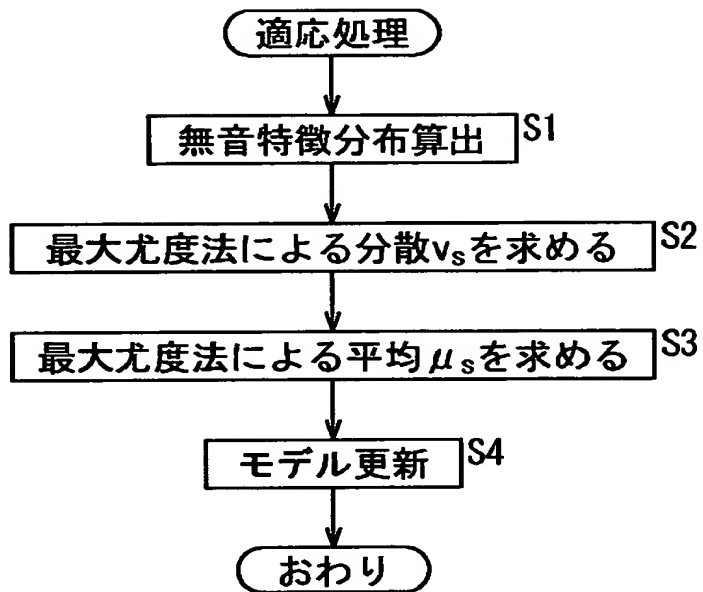


最大尤度法によるモデル適応

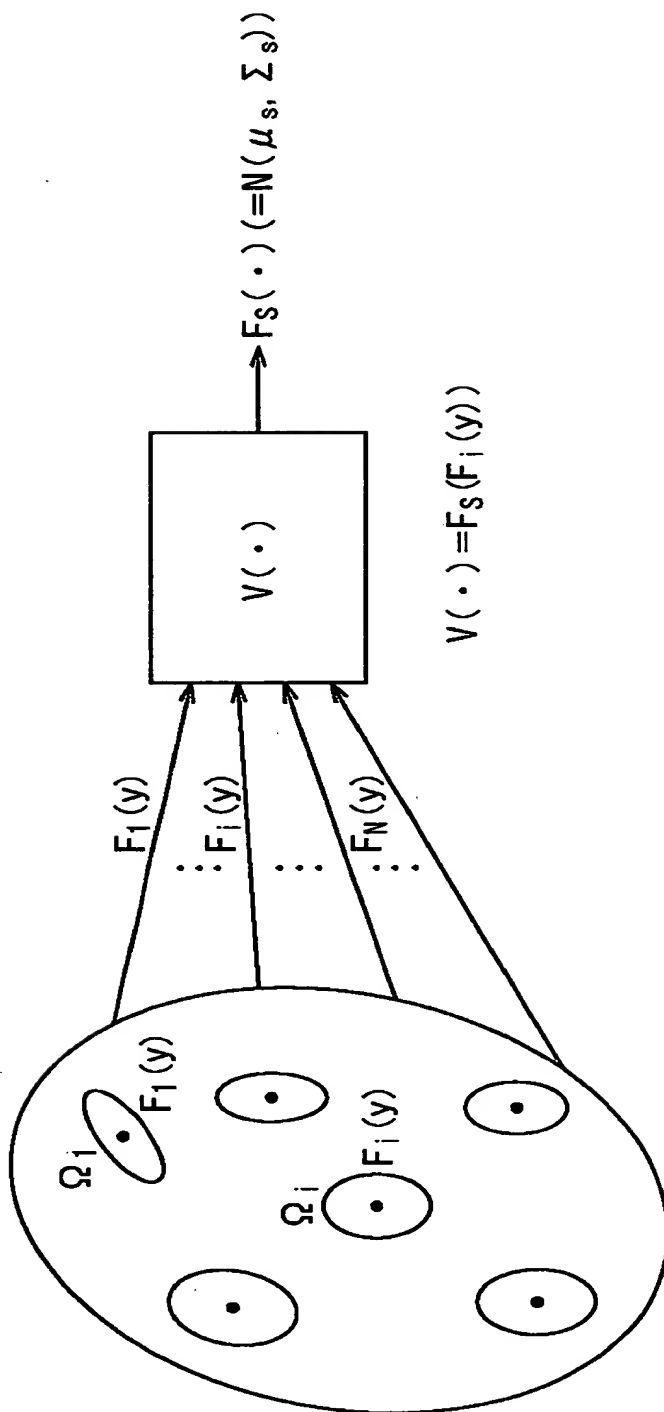
【図9】



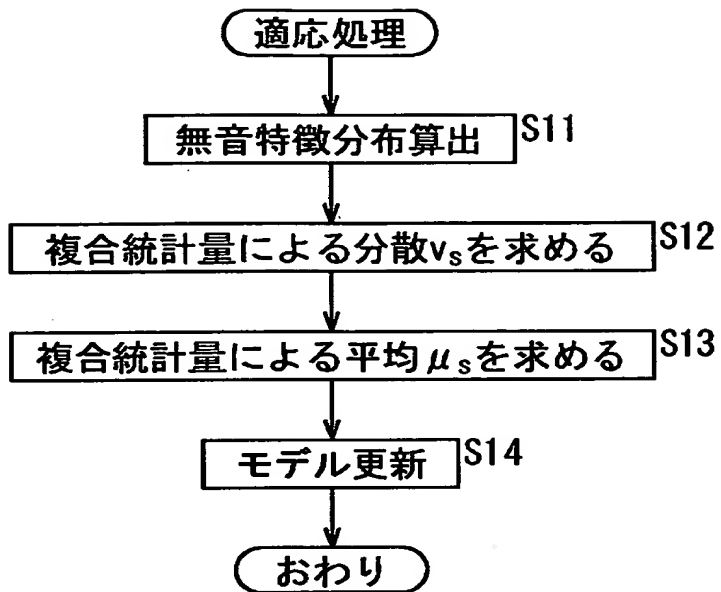
【図 1 0】



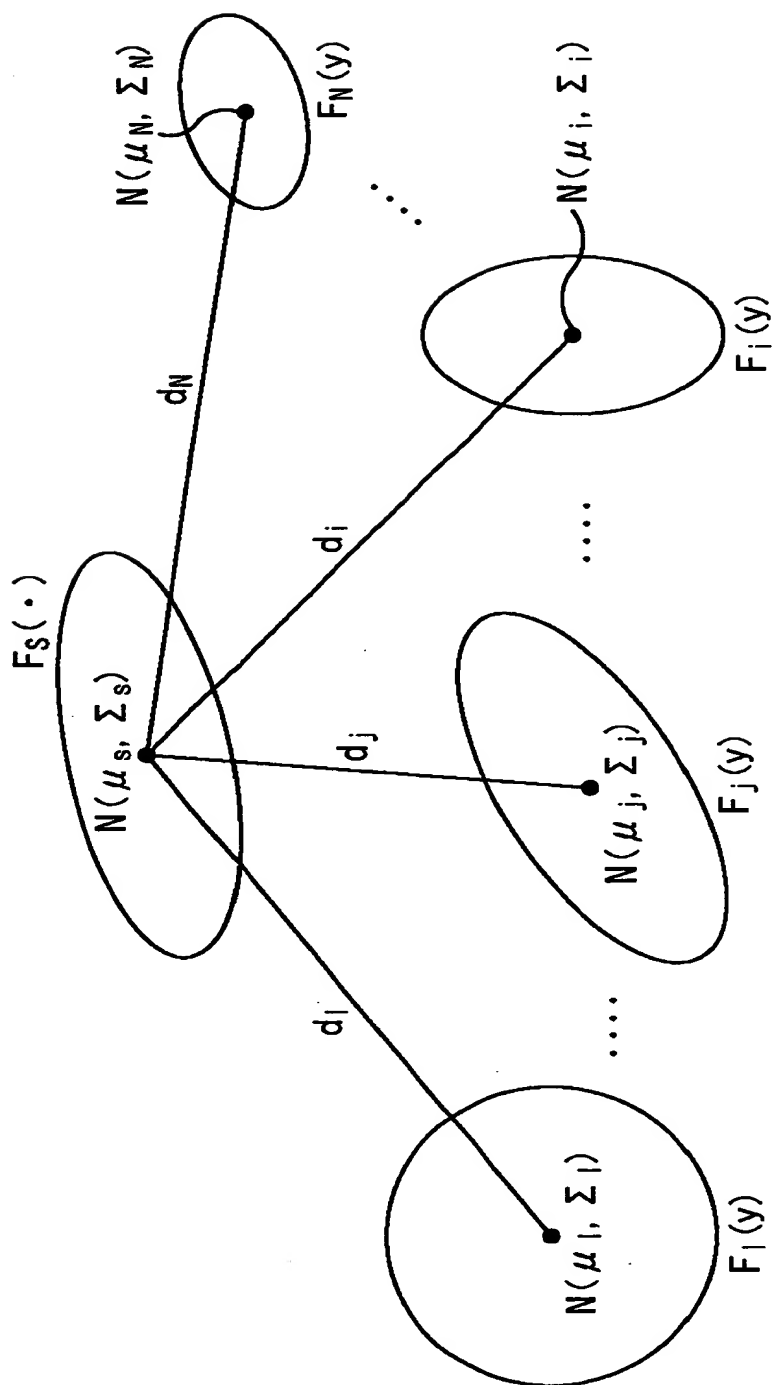
【図 11】



【図 1 2】

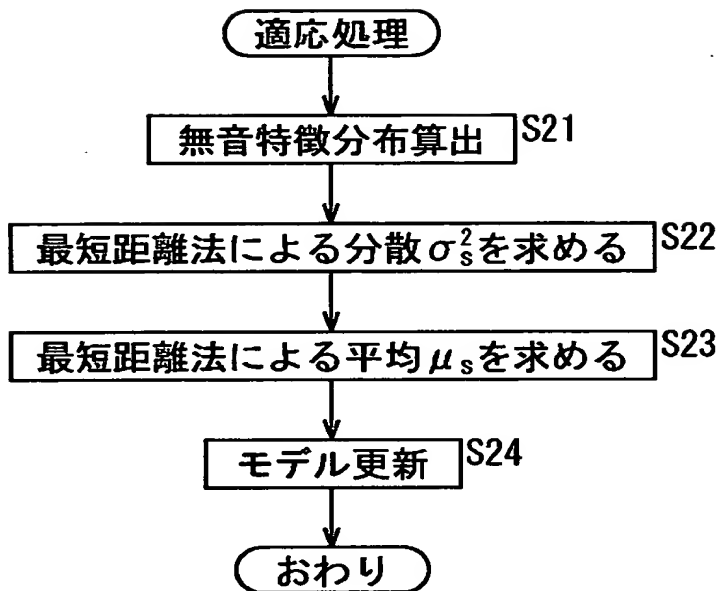


【図13】

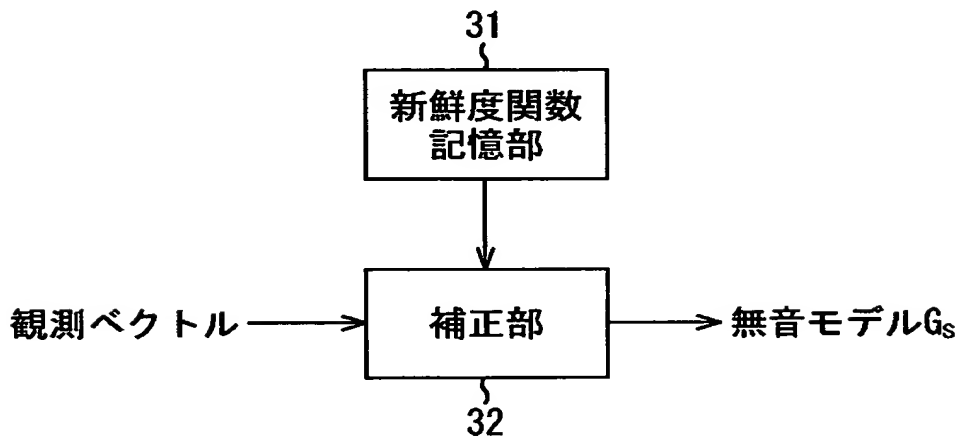


最短距離法によるモデル適応

【図14】

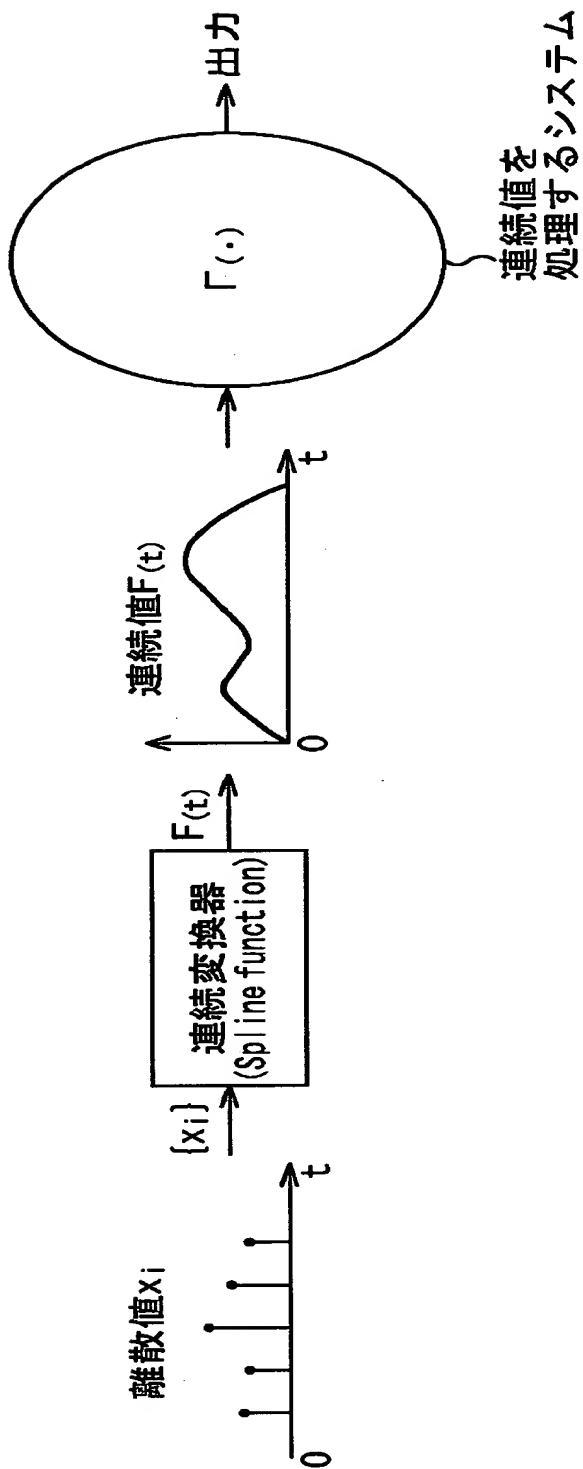


【図15】

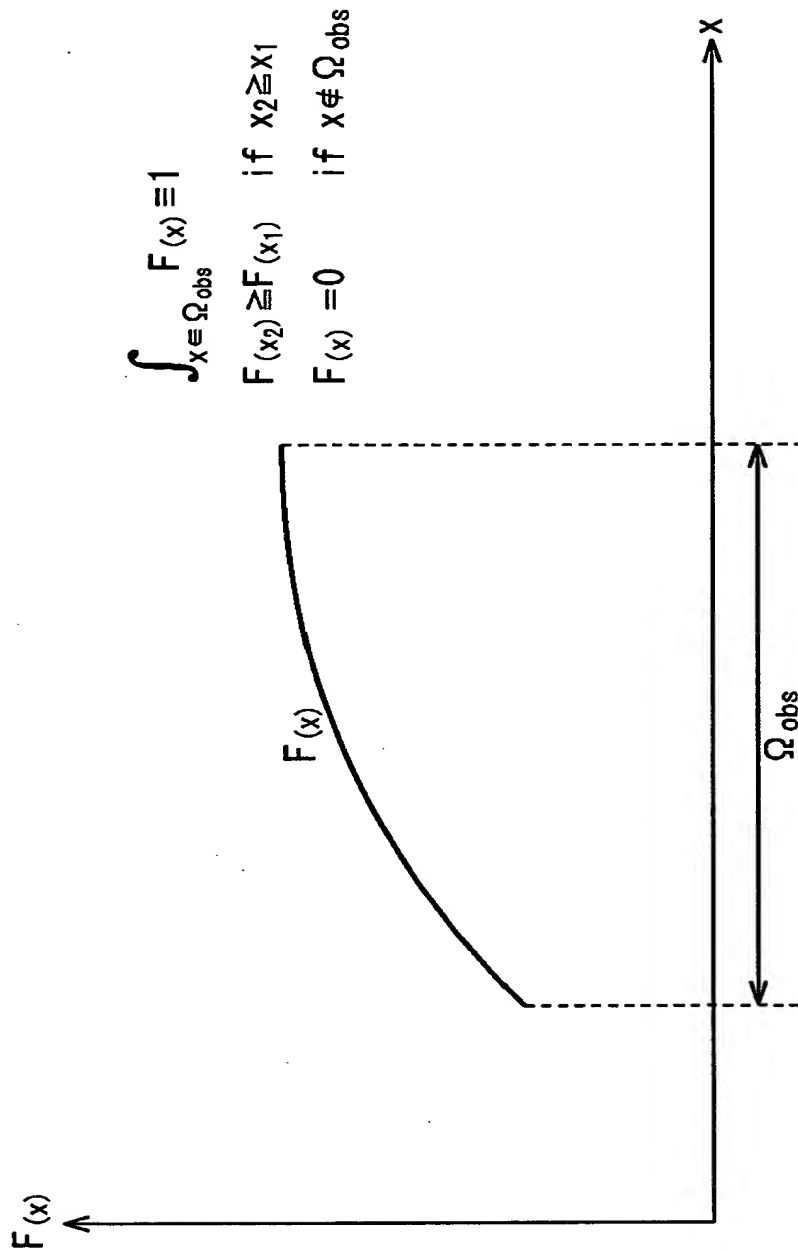


無音音響モデル補正部 7

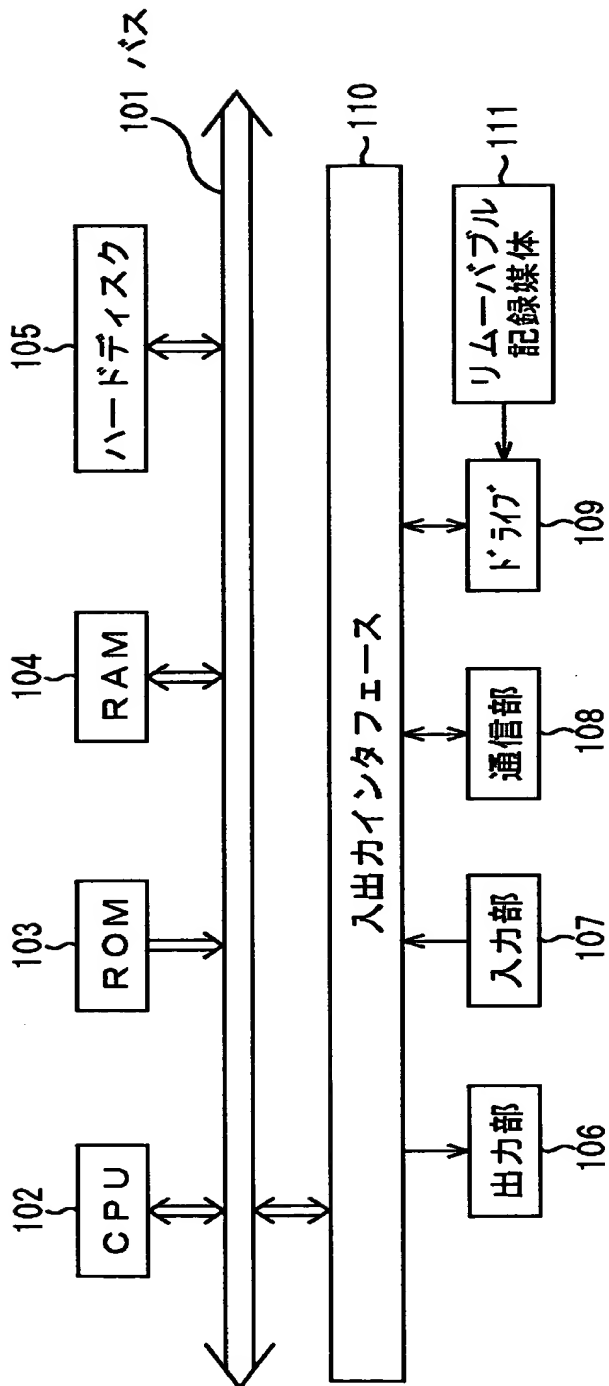
【図16】



【図 17】



【図18】



コンピュータ

【書類名】 要約書

【要約】

【課題】 認識性能を向上させる。

【解決手段】 無音音響モデル補正部 7 は、無音の状態を表す音響モデルである無音音響モデルの適応を、音声認識の対象とする音声認識区間の直前の区間で観測される入力データを用い、最尤法、複合統計量、または最短距離－最大分離定理によって行う。

【選択図】 図 1

特2000-264035

出 願 人 履 歴 情 報

識別番号 [000002185]

1. 変更年月日	1990年 8月30日
[変更理由]	新規登録
住 所	東京都品川区北品川6丁目7番35号
氏 名	ソニー株式会社